AFGL-TR-77-0169

REFINEMENTS IN THE COMBINED ADJUSTMENT OF SATELLITE ALTIMETRY AND GRAVITY ANOMALY DATA

Georges Blaha

DBA Systems, Inc.
Post Office Drawer 550
Melbourne, Florida 32901

July 1977

Final Report
(Period covered 29 September 1975 - 31 August 1977)

Approved for public release; distribution unlimited.

Prepared for:
DEFENSE MAPPING AGENCY
Building 56, U. S. Naval Observatory
Washington, D.C. 20305

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM A.F.B., MASSACHUSETTS 01731

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFGL - TR - 77 - 0169

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): REFINEMENTS IN THE COMBINED ADJUSTMENT OF SATELLITE ALTIMETRY AND GRAVITY ANOMALY DATA

5. TYPE OF REPORT & PERIOD COVERED: Scientific - Final Report. 29 Sept 75 - 31 July 77

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): Georges Blaha

8. CONTRACT OR GRANT NUMBER(s): F19628-76-C-0089

9. PERFORMING ORGANIZATION NAME AND ADDRESS
DBA Systems, Inc.
Post Office Drawer 550
Melbourne, Florida 32901

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 63701B 32050301 03

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Geophysics Laboratory
Hanscom AFB, Massachusetts 01731
Contract Monitor: George Hadgigeorge/LWG

12. REPORT DATE: 12 July 1977

13. NUMBER OF PAGES: 256

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 257p.

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Geoid	Distrubing potential
Reference ellipsoid	Spherical harmonics
Mean earth ellipsoid	Geoid undulation
Potential	Gravity anomaly

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

One of the most important objectives accomplished during the present analysis has been the upgrading of the AFGL computer program SAGG (Satellite Altimetry and Ground Gravity). This program serves in the determination of the global geoid and the earth's gravity field, based on the combination of satellite altimetry observations and gravity anomalies. A typical feature of SAGG is the simultaneous recovery of both the orbital parameters and the spherical harmonic potential coefficients. The short arc adjustment mode makes these determinations possible without the requirement of highly pre-

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

405 059

Unclassified (over)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

cise reference orbits. Perhaps the most important refinement in SAGG has been the differentiation of the radial distance to a sub-satellite geoidal point with respect to the state vector parameters. A practical benefit of this feature is faster convergence in the adjustment, which has removed the need for iterated solutions.

The new version of SAGG has been used in a combined adjustment of real data, in conjunction with a (14, 14) geopotential model. A comparison of internal precision has demonstrated the beneficial effect of adding altimetry data to the existing body of gravity anomaly data. However, adding this relatively new and completely independent source to a more traditional type of data in a combined adjustment is beneficial also for other reasons, such as increasing the external reliability of the results. The recovered geoid over most of the globe shows good agreement with gravimetric geoids obtained from independent sources. This is especially true of the areas covered by the GEOS-3 satellite when compared with the earlier reported results of the AFGL computer program SARRA (Short Arc Reduction of Radar Altimetry).

As a separate task, a theoretical study has been performed in comparing the spherical harmonic representation of the potential with the spheroidal harmonic representation. The outcome of this study has indicated that truncated expressions in any coordinate system other than the spherical (e.g., in a spheroidal coordinate system) would be essentially equivalent to using spherical harmonics, and they would be far less convenient.

The last, independent part of the present effort has been devoted to an analysis of the point mass technique. The purpose of introducing the point mass parameters into the adjustment is to add fine structure to a geopotential model that is based on spherical harmonic coefficients. The main attractive attributes of such an approach are its economy and its flexibility. In particular, in areas where an accurate geoid determination is not required or cannot be accomplished due to the lack of data, no point masses would be introduced; the geoid in such areas would accordingly be described by potential coefficients alone. This means that on the whole, the number of parameters could be kept relatively small. The point masses would be employed in areas where, for various reasons, a more detailed knowledge of geoid features may be required. The point mass technique has an immediate appeal not only for its relative simplicity and its practical applicability, but also because it is founded on a plausible physical concept.

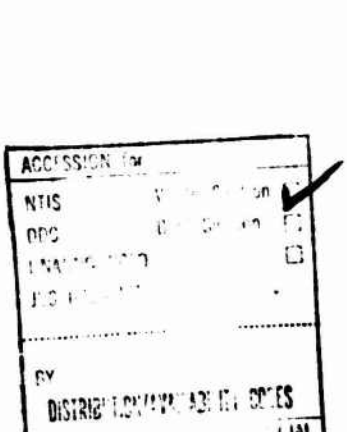


TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The past few years have seen the conception and development of the computer program SAGG (Satellite Altimetry and Ground Gravity). This program fulfills, at least partially, the needs of the Air Force Geophysics Laboratory (AFGL) for the determination of the global geoid and the earth's gravity field, based on the combination of satellite altimetry observations and gravity anomalies. A typical feature of SAGG is the simultaneous recovery of both the orbital parameters and the spherical harmonic potential coefficients describing the gravity field. The short arc adjustment mode makes these determinations possible without the requirement of highly precise reference orbits.

A more general insight into the concept of a short arc is offered in Section 2.1. As a practical outcome of this relatively new approach, the utilization of satellite altimetry is not dependent on intensive external tracking. It can be visualized along the following lines:

(a) Reference orbits generally accurate e.g. to 20 meters (rms) are initially established from routine global tracking (VHF Doppler tracking alone is sufficient to establish suitable reference orbits).

(b) The reference orbits are divided into a large number of subarcs situated over oceanic regions; such subarcs are limited in length to at most 1/4 revolution and average about 1/6 revolution.

(c) Each subarc is treated as an independent orbit with the epoch at midarc having a state vector subject to a-priori weighting consistent with the quality of the reference orbit.

(d) For each subarc, observation equations arising from satellite altimetry are introduced (such equations, of course, introduce a model representing the oceanic geoid).

(e) The adjustment attempts the simultaneous recovery of all geoidal parameters along with revised estimates of the state vectors of all orbital arcs (which may number in the thousands for dense global coverage).

The current analysis has been aimed at improving the theory and software related to the treatment of the satellite altimetry and gravity-type data. It is thus a natural continuation of efforts reported in [Brown, 1973] and [Blaha, 1975]. Many of the theoretical findings and improvements have served directly for modification of the previous version of SAGG. The new generation SAGG has the important practical feature that it can handle the real satellite altimetry data alone, the real gravity anomaly data alone, and the combination of the two types of data (a least squares adjustment process is employed).

The most important improvement to the satellite altimetry part of SAGG is described in Chapter 2. Its main outcome is the refinement in the partial derivatives entering the observation equations. It is shown that the radial distance (r) from the coordinate origin to the satellite should be differentiated not only with respect to the potential coefficients, but also with respect to the state vector components (the orbital parameters). A practical benefit of this refinement is a faster convergence in the adjustment, which has removed the need for iterated solutions.

Another important improvement related to satellite altimetry and implemented in SAGG is described in Chapter 3. A new weighting scheme is developed which takes into account the realistic error ellipsoids associated with the state vector parameters. These ellipsoids are invariably elongated along the in-track direction and are far from having a spherical shape as it had been assumed in the previous versions of SAGG. Further improvements which result in more efficient algorithms are mentioned in Chapter 4, in conjunction with extensive and independent testing of the orbital part of SAGG. The explicit form of the satellite altimetry observation equations may also be found in this chapter.

The gravity anomaly model utilized in a separate part of SAGG is treated in Chapter 5. A detailed study and a careful qualitative evaluation of the model is made before it is selected for the real data reductions. In the theoretical derivations, it is assumed that the reference ellipsoid has the same mass as the actual earth and the same potential as the geoid. This assumption has been adhered to throughout most of the present analysis. It is well justified whenever the reference ellipsoid is identified with the mean earth ellipsoid which shares with the actual earth four defining parameters (i.e., two additional parameters besides the mass and the potential). Since the mean earth ellipsoid is the best global approximation of the earth by an ellipsoid of revolution, the values of the mass and potential as supplied by this ellipsoid can indeed be considered fixed in a global type of adjustment.

The above statement can be represented mathematically as follows. If W_o is the potential of the geoid and U_o is the potential of the reference ellipsoid (assumed to be the mean earth ellipsoid), then one can write

$$W_o \equiv U_o \quad , \quad dW_o \equiv 0 \quad , \tag{1.1}$$

where dW_o is the correction to the orginal value W_o. Similarly, if M is the earth's mass and M' is the mass of the same reference ellipsoid, it is considered that

$$M \equiv M' \quad , \quad dM \equiv 0 \quad , \tag{1.2}$$

where dM is the correction to the original value M. It will prove useful to replace the quantity W_o by an auxillary quantity r_o, such that

$$r_o = kM / W_o \quad , \tag{1.3a}$$

where k is the gravitational constant. In the same way, one would have

$$r_o^* = kM' / U_o \quad . \tag{1.3b}$$

Due to (1.1) and (1.2), an equivalent relation is obtained:

$$r_o \equiv r_o^* \quad , \quad dr_o \equiv 0 \quad , \tag{1.4}$$

where dr_o is the correction to the original value r_o. Equations (1.2) and (1.1) or, equivalently, (1.4) are used throughout this analysis; the only exceptions occur in Chapter 2 (the existence of dr_o is noticed in some very general expressions) and in Chapter 7, where r_o is considered an adjustable (weighted) parameter.

A combined adjustment is treated in Chapter 6, which also contains a description of some new, desirable features of SAGG, such as the data handling capabilities or the various options for weighting the potential coefficients. In Chapter 7, a real data adjustment accomplished with the latest version of SAGG is described and its outcome is presented graphically and verbally. Both the satellite altimetry and gravity anomalies are part of this adjustment process. The altimetry data processed by SAGG has been gathered by the GEOS-3 satellite over adjacent areas of the Indian and South Pacific Oceans and a portion of the North Atlantic; the gravity anomaly data is represented by mean anomalies from over 2200 $1^0 \times 1^0$ geographic blocks.

Chapter 8 and Chapter 9 stand apart from all the other chapters of the present analysis. They each contain an independent theoretical study and result in several recommendations; however, they are not accompanied by any software development. The topic of Chapter 8 is a comparison of two different expansions of the potential, the spherical harmonics and the spheroidal harmonics. Chapter 9 is devoted to an analysis of the point mass technique which can serve in adding detail to a geopotential model based on the spherical harmonic coefficients. The derivations are compatible with the expressions (1.1) and (1.2).

A few most important results and recommendations are summarized in Chapter 10, which is also the last chapter in the body of the present study. Two separate topics are treated in Appendix 1 and Appendix 2. The formulas from Appendix 1, dealing with the geometry of satellite altimetry, are referenced in several instances, for example in Chapter 2 or in Chapter 4.

It is noteworthy that most of the chapters, as well as both appendices, are presented essentially in a self-contained manner. An advantage of this format is that they can be read independently without undue damage to a good understanding of the material. A typical example in this respect is Chapter 2, which, with a few minor changes, has been presented separately as [Blaha, 1977].

2. REFINEMENT OF THE SHORT ARC SATELLITE ALTIMETRY ADJUSTMENT MODEL

2.1 Background

The highlight of the approach to satellite altimetry initiated by Brown [1973] is the use of the short arc adjustment mode. This feature makes a determination of the geoid surface possible without the requirement of highly precise reference orbits. In fact, approximate values of the state vector parameters with some fairly large a-priori standard deviations are sufficient. The observations gathered over the oceanic regions represent the distances from a satellite to the geoid. For many purposes, the geoid surface is assumed to coincide with the surface of oceans. For a good determination of the geoid everywhere, it is necessary to combine satellite altimetry data with some other kind of data gathered over the continental regions, such as gravity observations.

In this treatment, we confine our attention to the satellite altimetry model. The parameters are divided into two basic groups: 1) the earth potential coefficients which are the conventional C's and S's of the spherical harmonic expansion including the "central term" r_0, and 2) six state vector components for each short arc. The parameters in the first group may be considered completely free to adjust, except for the five "forbidden" coefficients, $C_{10}, C_{11}, S_{11}, C_{21}, S_{21}$, which are usually set to zero and held at that value; this implies that the origin of the Cartesian coordinate system is identical with the earth's center of mass and that the z-axis coincides with the earth's axis of rotation. The

parameters in the second group are assumed to be supplied from an independent source together with their a-priori error characteristics. They may be given in one inertial coordinate system for all arcs, or in a chosen inertial coordinate system which may vary from arc to arc and may coincide, for convenience, with the earth fixed (E.F.) coordinate system at the epoch time, or in the E.F. system itself. We choose this last possibility which leads to a whole adjustment being performed in the E.F. system.

A brief discussion of the pertinent short arc characteristics is in order. As pointed out in [Brown, 1967], the rationale for the short arc approach is that orbital errors resulting from the enforcement of a reasonably accurate set of potential coefficients truncated at fairly low degree and order can under proper circumstances be accommodated by slight adjustments of the six state vector parameters. For example, in the case of a nearly circular orbit at an altitude of 1000 km, the integration of an arc of 600 seconds using enforced potential coefficients (constants) truncated at (n,m) = (4,4) can result in errors at the extremities of the arc approaching 1 to 2 meters when the state vector at mid-arc is held to its correct value. However, when the state vector is allowed appropriate freedom to adjust, the error in the resulting arc may be reduced to a level of a few centimeters. In the general context of the short arc method, then, a "short arc" is simply one that is sufficiently short so that positional errors attributable to the errors in enforced potential coefficients (combined with errors of truncation) are "sufficiently well" accommodated over the arc by an adjustment of the state vector. What constitutes

"sufficiently well" depends on the character of the particular problem. By virtue of such considerations, it can be said that the short arc method, by definition, does not require that specific attention be paid to errors in potential coefficients insofar as each satellite arc is concerned. This is compatible with the functional relationship appearing in equation (43) of [Brown, 1973]. In particular, the coordinates of a satellite point on a short arc are expressed as independent of the adjustable potential coefficients.

The state vector components for a short arc adjustment have often associated with them some standard errors which serve subsequently for weighting purposes. Weighting must be independent for each arc as documented in [Blaha, 1975], pages 15 and 17. As is known, weights for orbital parameters, weak though they may be, are required in a short arc satellite altimetry adjustment; without them the determination of all the six state vector parameters would be all but impossible since altimeter measurements contain little or no information about the spatial orientation of the orbital plane. A discussion to this effect may be found in [Brown, 1973], pages 17 and 18.

2.2 General Outline

In the forthcoming discussion, the radial distance to a satellite point (S) will be denoted by the letter R and the radial distance to the geoid, by the letter r, both symbols representing the adjusted quantities that fulfill exactly a prescribed mathematical model. The approximate values of these quantities based on some initial values of parameters will be denoted as R^0 and r^0. The adjusted and observed satellite altimetry values will be denoted as H and H^b, respectively. These notations will serve only when forming the "discrepancy" terms (sometimes called constant terms) in observation equations; in other instances, especially when forming the partial derivatives, the indices will be simply dropped. Thus the state vector components for an arc are symbolized by $X,Y,Z,\dot{X},\dot{Y},\dot{Z}$ and the coordinates of S, a point on this arc, by X_S, Y_S, Z_S (all in the E.F. system). The differentiation of R with respect to the state vector components is done via the chain rule applied to matrices as follows:

$$\frac{\partial R}{\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z})} = \frac{\partial R}{\partial(X_S,Y_S,Z_S)} \frac{\partial(X_S,Y_S,Z_S)}{\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z})} .$$

Since $R = (X_S^2 + Y_S^2 + Z_S^2)^{\frac{1}{2}}$, the first matrix on the right-hand side is composed of the three directional cosines, and the second (3x6) matrix, denoted $\tilde{M}_1$, is given as

$$\tilde{M}_1 = \frac{\partial(X_S,Y_S,Z_S)}{\partial(x_S,y_S,z_S)} \frac{\partial(x_S,y_S,z_S)}{\partial(x,y,z,\dot{x},\dot{y},\dot{z})} \frac{\partial(x,y,z,\dot{x},\dot{y},\dot{z})}{\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z})} ;$$

here, the lower case letters are equivalent to their counterparts (capital letters), except that they refer to a chosen inertial system. The first (3x3) matrix on the right-hand side of the last equation represents a rotation and the second (3x6) matrix consists of the first three rows of the matrizant. For our purpose, $\tilde{M}_1$ will not be expressed explicitly. For the partial derivatives of R we now write

$$\partial R/\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z}) = \left[X_s/R,\ Y_s/R,\ Z_s/R\right]\tilde{M}_1. \tag{2.1}$$

In view of the previous remarks concerning the short arc approach, the differentiation of R with respect to C's and S's will not be considered.

The radial distance to the geoid is given in [Blaha, 1975], equation (46a), as

$$r = r_o\left[1 + \sum_{n=2}^{\infty}(a/r)^n \sum_{m=0}^{n}(C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)P_{nm}(\sin\bar{\phi})\right]$$

$$+\tfrac{1}{2}\omega^2 r_o r^3 \cos^2\bar{\phi}/(kM), \tag{2.2}$$

where $\bar{\phi}$ and λ are the geocentric latitude and longitude, a is the earth's equatorial radius, ω is the angular velocity of earth's rotation, and kM is the product of the gravitational constant and the earth's mass; r_o, C_{nm} and S_{nm} were already labeled as r_o, C's and S's, and $P_{nm}(\sin\bar{\phi})$ are the conventional Legendre functions. The series expansion in (2.2) is truncated at some suitable n = N: this also constitutes a reason why less concern has been attached recently to the convergence problems associated with such series (see e.g. [Hotine, 1969], page 173, [Needham, 1970],

pages 46 and 47, or [Rapp, 1972], page 20). The parameter r_0 deserves further attention. It corresponds to the radius of a fictitious spherical earth having the same potential as the geoid; namely,

$$r_0 = kM/W_0,$$

where W_0 is the gravity potential of the geoid. To obtain a starting value of r_0 in case it is subject to adjustment, one could use U_0, the potential of the mean earth ellipsoid in the Geodetic Reference System (GRS) 1967, given as

$$U_0 = 6.2637030523 \times 10^9 \text{ gal} \times \text{m};$$

the value of kM in this system is

$$kM = 3.98603 \times 10^{14} \text{ m}^3/\text{sec}^2.$$

Accordingly, we could adopt for r_0 (it represents essentially the scale of the earth):

$$r_0 = 6.363696 \times 10^6 \text{ m}.$$

In what follows, we shall consider two mathematical models of satellite altimetry. The first model will be very elementary, acceptable nevertheless for error propagation purposes as well as for some preliminary data reductions. The second model will be more rigorous, not only with regard to the discrepancy terms but also with regard to partial derivatives appearing in observation equations. In this model, a correction will be given for individual discrepancy terms and another

correction will be derived for partial derivatives with respect to the state vector. It will be apparent that both these corrections are extremely simple to apply which makes it an easy matter to upgrade computer programs conceived in terms of the elementary model. The purpose of this refinement is to make the observation equations as rigorous as possible, compatible with a specified mathematical model. Herein we treat the short arc satellite altimetry adjustment model, although similar considerations could apply also in other cases. Besides offering a deeper insight into this model, the presented analysis will also benefit the practitioner. It will result in faster convergence which, in some cases, may remove the need for iterated solutions altogether.

2.3 Elementary Model

The mathematical model we are considering is depicted in Figure 2.1. It also appeared in [Blaha, 1975], Section 5.1, and it reads

$$H = R - r. \tag{2.3}$$

This would imply that H is measured along the position vector to a satellite, while in reality it is measured perpendicular to the geoid. Due to the flattening, such an approximation could introduce an error of several meters in magnitude.

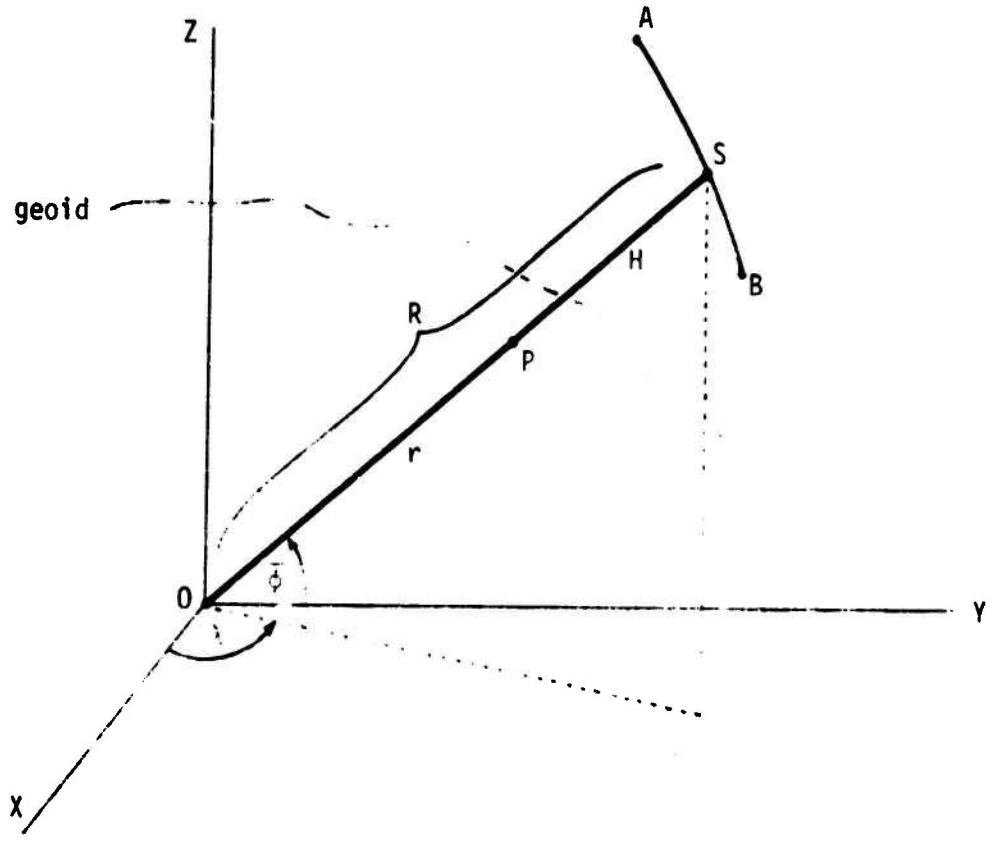


FIGURE 2.1

Simplified mathematical model. Satellite altitude H is measured along the position vector to a satellite (S) on the short arc AB.

As far as the differentiation of H in (2.3) with respect to parameters is concerned, a simplified approach would correspond to differentiating R with respect to the state vector components (X,Y,Z, $\dot{X}$.$\dot{Y}$,$\dot{Z}$) and to differentiating r with respect to the potential coefficients (r_0,C's and S's). The shortcoming of this approach is the fact that r in (2.2) is also a function of the geocentric latitude and longitude, which in turn depend on the state vector components. Thus, to a certain extent, r is also a function of these components and should be differentiated accordingly. Here again, the bulk of the correction to be applied later to the partial derivatives is due to the flattening. Under the present simplifications, the observation equation corresponding to (2.3) is

(2.4)

$$V = [X_s/R,\ Y_s/R,\ Z_s/R]\tilde{M}_1 \begin{bmatrix} dX \\ dY \\ dZ \\ d\dot{X} \\ d\dot{Y} \\ d\dot{Z} \end{bmatrix} - [\partial r/\partial(r_0, C's, S's)] \begin{bmatrix} dr_0 \\ dC's \\ dS's \end{bmatrix} + [(R^0 - r^0) - H^b],$$

where (2.1) has been taken into account and where the last term had previously been referred to as "discrepancy" term.

2.4 More Rigorous Model

Assuming that H is measured perpendicularly to the geoid, the points O,P,S in Figure 2.1 would no longer be on a straight line. In general, they are not lying in a meridian plane (see Figure 2.2). In this study, we shall consider only the corrections to the elementary model due to the flattening, which will nevertheless take care of the bulk of the discrepancy between such a model and the real world. Replacing the geoid by a geocentric ellipsoid for this purpose, we state that the plane of OPS is now a meridian plane and that the correction to the model equation (2.3) depends on the latitude but not on the longitude.

We may now write the mathematical model of satellite altimetry as

$$H = R - r + d, \tag{2.5}$$

where d is the above mentioned correction. We note that H is the distance from a satellite to the geoid (and not to the ellipsoid) and the ellipsoid serves only to approximate the direction along which this distance is measured in order to compute d. The error in d caused by this approximation would amount to less than 4 cm even if the deflection of the vertical were 1' (for simplicity, we usually take H ≈ 1000 km). Considering the overall precision of satellite altimetry, such errors are negligible. It can be shown that

$$d \simeq \tfrac{1}{2}\, r\, \delta^2\, (r+H)/H$$

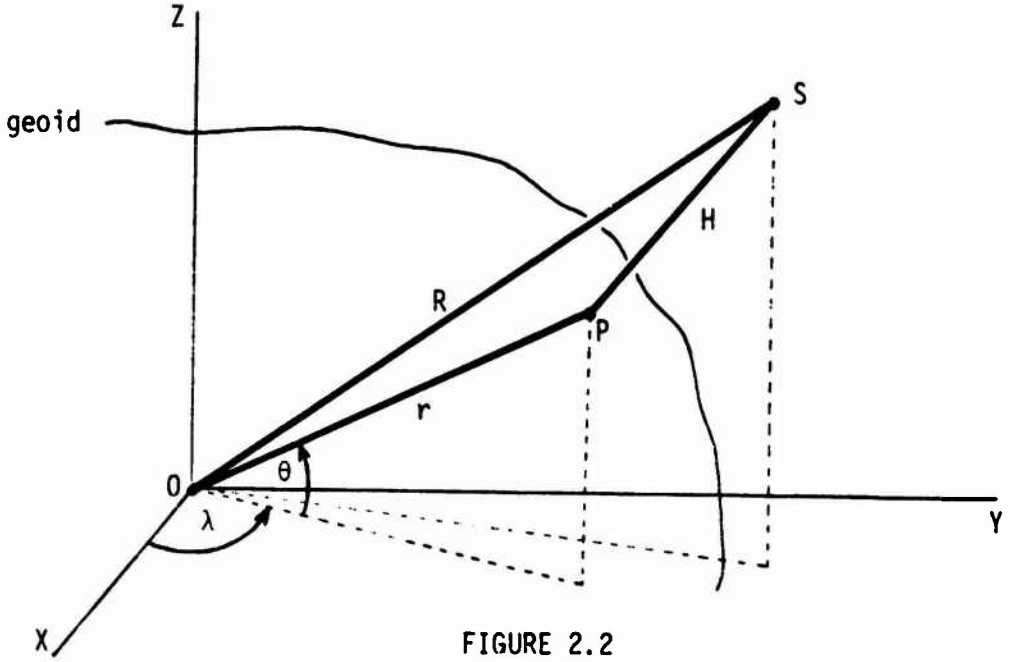


FIGURE 2.2

Realistic situation. Satellite altitude H is measured perpendicular to the geoid.

(see equation A1.20 in Appendix 1), where

$$\delta = \bar{\phi} - \theta,$$

$\bar{\phi}$ being the geocentric latitude of S and θ representing the geocentric latitude of P as in Figure 2.2. Adopting henceforth the values

$$e^2 \simeq 0.006695 \qquad (2.6)$$

for the reference ellipsoid (rounded value of the mean earth ellipsoid in the GRS 1967) and

$$H \simeq 1000\ \text{km} \qquad (2.7)$$

for the satellite altitude (all the formulas will be nearly linear in H for relatively low satellites), we find that δ will not exceed 1.5'. The

analytical expression for d may also be given to a good approximation as (see equation A1.21)

$$d \simeq (1/8)\, e^4\, r_0[H/(r_0+H)]\sin^2 2\theta; \qquad (2.8)$$

r_0 represents again a scale, in Appendix 1 given the numerical value of the mean earth's radius,

$$r_0 \simeq 6371 \text{ km}. \qquad (2.9)$$

We thus find that

$$d \simeq 4.9\text{m} \sin^2 2\theta,$$

considered in conjunction with (2.7). In this case, the correction d, always positive, amounts to 4.9m at the most (at 45^0 latitude).

We mention that a formula similar to our equations (2.5), (2.8) had previously been derived, in a different way, in [Gopalapillai, 1974]. However, if we set N=0 in (2.9) on page 10 of this reference, we recover our equation (2.5), except for the sign associated with the non-negative quantity d (thus, in the extreme case, instead of adding 4.9m one would be subtracting 4.9m). A similar consideration would apply with regard to the formula on page 11 of this reference. We submit that in our Figure 2.2, a positive quantity should be added to R-r in order to yield H, since from the geometry of the figure it follows that H>R-r.

It is clear that except for the above difference in signs, our formula for d does not represent a new result in itself. In fact, the

need for it would entirely disappear if, instead of (2.5), the vector equation

$$\vec{H} = \vec{R} - \vec{r}$$

were used to compute the discrepancy term. In particular, instead of the relation

$$OP = R - H + d, \tag{2.9'}$$

one could use the equation

$$OP = (X_P^2 + Y_P^2 + Z_P^2)^{\frac{1}{2}} \tag{2.9"}$$

in agreement with Figure 2.2, where

$$X_P = X_S - H \cos\phi \cos\lambda,$$
$$Y_P = Y_S - H \cos\phi \sin\lambda,$$
$$Z_P = Z_S - H \sin\phi;$$

the direction of $\vec{H}$ represented by ϕ,λ could again be sufficiently well approximated by solving for the sub-satellite ellipsoidal point. The coordinates of the sub-satellite geoidal point computed from the last three equations are needed in evaluating the discrepancy term, whether the approach with "d" (as in 2.9') or the more direct approach (as in 2.9") is used. In particular, the angle

$$\theta = \text{arc tg}[Z_P/(X_P^2 + Y_P^2)^{\frac{1}{2}}] \tag{2.9'''}$$

is required to compute r according to (2.2) since in the more rigorous model, $\bar{\phi}$ should be replaced by θ(the geocentric latitude of a geoidal point).

One might then be compelled to ask a perfectly legitimate question. What use has the formula for d if, in the given context, it accelerates the computations by only completely negligible amounts (equation 2.9' versus 2.9")? The answer is that the importance of this formula is more theoretical than practical. As we shall see, it serves in deriving the expressions for partial differentiation of the satellite altimetry mathematical model with respect to the parameters. We have seen that

$$H \neq R - r$$

in the more rigorous model; this being the case, how could we form the partial derivatives $\partial R/\partial$ parameters and $-\partial r/\partial$ parameters and claim, without a formal proof, that the observation equations may be built in this manner? However, with an explicit formula for d available, such differentiation can be justified by showing that the contribution of the partial derivatives $\partial d/\partial$ parameters is negligible. This problem will be addressed next. One will appreciate that if satellite altimetry measurements were performed on a planet whose flattening were much greater than that of the Earth, the formulas for partial differentiation would indeed change. The development leading to the partial derivatives $-\partial r/\partial$ parameters will then follow. Although r is principally a function of the potential coefficients, it becomes dependent on the orbital elements through the altimeter measurement which is considered perpendicular to the geoid or, to a good approximation, perpendicular to the ellipsoid.

The analytic differentiation of d with respect to the parameters is performed first considering the earth's potential coefficients (in this case represented only by r_0) and then considering the state vector components. For the current purpose, we approximate (2.8) further by writing

$$d \simeq (1/8)\, e^4 H \sin^2 2\theta (1-H/r_0),$$

from which it follows that

$$\partial d/\partial r_0 \simeq (1/8)\, e^4 (H/r_0)^2 \sin^2 2\theta,$$

or, with (2.6), (2.7), and (2.9),

$$\partial d/\partial r_0 \simeq 0.1 \times 10^{-6} \sin^2 2\theta.$$

Deducing from (2.2) that

$$\partial r/\partial r_0 \simeq 1, \tag{2.10}$$

we conclude that $\partial d/\partial r_0$ could affect at most the 7th significant digit in the partial differentiation with respect to r_0 and that it is therefore completely negligible. In fact, if the magnitude of discrepancy terms were tens or even hundreds of meters, the sufficient, comparable number of reliable significant digits in partial derivatives would be three to five, provided of course that none of these digits were lost due to round-off errors accumulated during actual computations. On the other hand, we can write

$$\partial d/\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z}) = (\partial d/\partial\theta)[\partial\theta/\partial(X_S,Y_S,Z_S)]\tilde{M}_1, \tag{2.11}$$

where, from (2.8),

$$\partial d/\partial\theta = \tfrac{1}{2}\, e^4\, r_0[H/(r_0+H)]\, \sin 2\theta \cos 2\theta. \qquad (2.12)$$

A relation similar to (2.11) will hold for (-r) replacing d and it will be called the "main correction term". If the partial derivatives appearing in (2.11) were not to be neglected, we would be dealing with a total correction whose first factor would be

$$-\partial r/\partial\theta + \partial d/\partial\theta \simeq \tfrac{1}{2}\, a\, e^2\, \sin 2\theta \{1 + e^2[H/(r_0+H)] \cos 2\theta\}, \qquad (2.13)$$

where we have written "a" for "r_0" in (2.12) and where the expression for $-\partial r/\partial\theta$ will be justified later. Upon substituting the usual numerical values in the second term inside the brackets, which represents the contribution of $\partial d/\partial\theta$, we find this term to be always smaller than 0.00090. This is certainly negligible compared to unity if we consider that the ellipsoidal approximation in this case leads to errors much greater than 0.09% of the "main correction term". Accordingly, we shall discard the partial derivatives of d with respect to any parameters.

This brings us to the "main correction term", the only important correction to be included in the process of partial differentiation in the present model. From (2.5) we see that it is obtained by differentiating (-r) with respect to the state vector components. Using again the chain rule in matrix notation, this will be done as follows:

$$-\partial r/\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z}) = -(\partial r/\partial\phi)(\partial\phi/\partial\theta)(\partial\theta/\partial\bar{\phi})[\partial\bar{\phi}/\partial(X_s,Y_s,Z_s)]\ddot{M}_1. \qquad (2.14)$$

The various quantities needed in this task are shown in Figure 2.3. For the point P on the ellipsoid, we have

$$r^2 = [N \cos\phi]^2 + [N(1-e^2) \sin\phi]^2,$$

where

$$N = a(1-e^2 \sin^2 \phi)^{-\frac{1}{2}},$$

and

$$r = a(1-e^2 \sin^2 \phi)^{-\frac{1}{2}} \ [1-e^2(2-e^2) \sin^2 \phi]^{\frac{1}{2}}. \qquad (2.15)$$

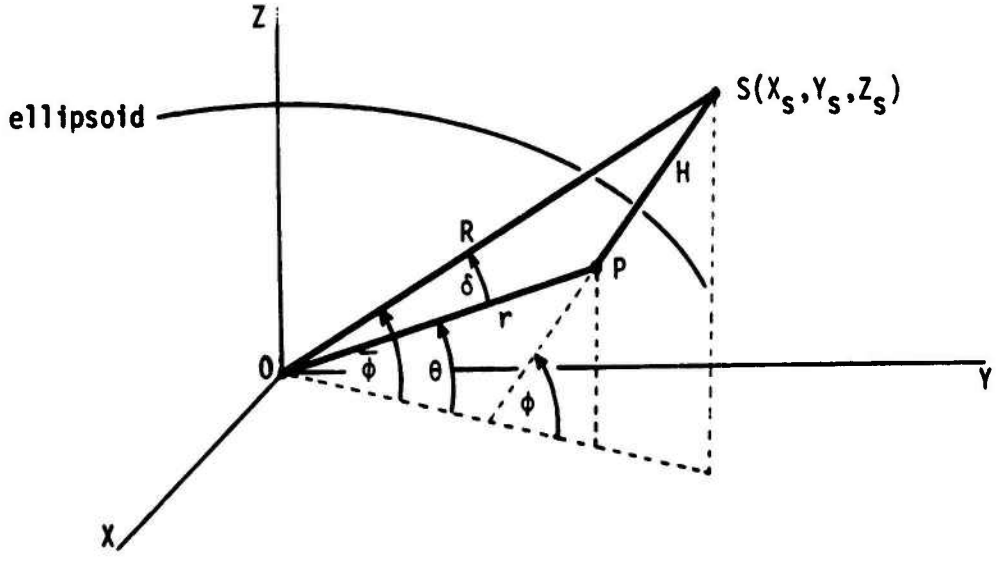


FIGURE 2.3

Ellipsoidal approximation of the geoid in view of partial differentiation of (-r) with respect to the state vector components

Differentiating the last equation with respect to ϕ, we find

$$\partial r/\partial\phi = -M\, e^2 \sin\phi \cos\phi/[1-e^2(2-e^2)\sin^2\phi]^{\frac{1}{2}},$$

where

$$M = a(1-e^2)/(1-e^2\sin^2\phi)^{\frac{3}{2}} .$$

Neglecting the terms with e^4, e^6, etc., we have

$$-\partial r/\partial\phi \simeq a\, e^2 \sin\phi \cos\phi. \qquad (2.15')$$

From the standard formula

$$\operatorname{tg}\theta = (1-e^2)\operatorname{tg}\phi$$

we obtain

$$\partial\phi/\partial\theta \simeq 1/(1-e^2\cos 2\theta)$$

and

$$\partial\phi/\partial\theta \simeq 1 + e^2\cos 2\theta, \qquad (2.16)$$

neglecting again e^4, etc. At this point we present the relation used in (2.13), namely

$$\begin{aligned} -\partial r/\partial\theta &= -(\partial r/\partial\phi)(\partial\phi/\partial\theta) \\ &\simeq a\, e^2 \sin\phi\cos\phi \\ &\simeq \tfrac{1}{2} a\, e^2 \sin 2\theta. \end{aligned}$$

From Figure 2.3 we read

$$\bar{\phi} = \theta + \delta.$$

In (A1.19) it is shown that

$$\delta \simeq \tfrac{1}{2}\, e^2[H/(r_0+H)]\, \sin 2\theta,$$

which in fact served in deriving (2.8); thus

$$\partial\bar{\phi}/\partial\theta \simeq 1+e^2[H/(r_0+H)]\, \cos 2\theta. \tag{2.17}$$

If we now substitute (2.15'), (2.16), and (2.17) into (2.14) and if we neglect the terms with e^4, etc., we obtain the following result:

$$-\partial r/\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z}) \simeq a\, e^2\, \sin\bar{\phi}\, \cos\bar{\phi}\, [\partial\bar{\phi}/\partial(X_S,Y_S,Z_S)]\tilde{M}_1. \tag{2.18}$$

The next step consists in finding the partial derivatives of $\bar{\phi}$ with respect to X_S, Y_S, Z_S. First we express the relationship between the Cartesian and the spherical coordinates of the point S (see also Figure 2.3):

$$\left.\begin{aligned} X_S &= R\, \cos\bar{\phi}\, \cos\lambda, \\ Y_S &= R\, \cos\bar{\phi}\, \sin\lambda, \\ Z_S &= R\, \sin\bar{\phi}. \end{aligned}\right\} \tag{2.19}$$

Then we differentiate it in the following manner:

$$A \equiv \partial(X_S,Y_S,Z_S)/\partial(R,\bar{\phi},\lambda)$$

$$= \begin{bmatrix} \cos\bar{\phi}\, \cos\lambda & -R\, \sin\bar{\phi}\, \cos\lambda & -R\, \cos\bar{\phi}\, \sin\lambda \\ \cos\bar{\phi}\, \sin\lambda & -R\, \sin\bar{\phi}\, \sin\lambda & R\, \cos\bar{\phi}\, \cos\lambda \\ \sin\bar{\phi} & R\, \cos\bar{\phi} & 0 \end{bmatrix}.$$

By inverting A analytically, which is possible whenever $\bar{\phi} \neq \pm\frac{1}{2}\pi$ (the determinant is $-R^2 \cos\bar{\phi}$), we obtain

$$A^{-1} \equiv \partial(R,\bar{\phi},\lambda)/\partial(X_s,Y_s,Z_s)$$

$$= \begin{bmatrix} \cos\bar{\phi}\cos\lambda & \cos\bar{\phi}\sin\lambda & \sin\bar{\phi} \\ -\sin\bar{\phi}\cos\lambda/R & -\sin\bar{\phi}\sin\lambda/R & \cos\bar{\phi}/R \\ -\sin\lambda/(R\cos\bar{\phi}) & \cos\lambda/(R\cos\bar{\phi}) & 0 \end{bmatrix}. \qquad (2.20)$$

As a matter of verification, we have $A^{-1}A = I$. The second row of the matrix in (2.20) reads

$$\partial\bar{\phi}/\partial(X_s,Y_s,Z_s) = (1/R)[-\sin\bar{\phi}\cos\lambda,\ -\sin\bar{\phi}\sin\lambda,\ \cos\bar{\phi}]$$

which, when substituted into (2.18), yields (2.21)

$$-\partial r/\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z}) \simeq (a/R)\, e^2[-\cos\bar{\phi}\cos\lambda\sin^2\bar{\phi},\ -\cos\bar{\phi}\sin\lambda\sin^2\bar{\phi},\ \sin\bar{\phi}\cos^2\bar{\phi}]\tilde{M}_1$$

$$= (a/R)\, e^2[-(X_s/R)\sin^2\bar{\phi},\ -(Y_s/R)\sin^2\bar{\phi},\ (Z_s/R)\cos^2\bar{\phi}]\tilde{M}_1,$$

where (2.19) has been used.

The observation equation corresponding to the model represented by (2.5) is

$$V = [\partial R/\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z})]\begin{bmatrix} dX \\ dY \\ dZ \\ d\dot{X} \\ d\dot{Y} \\ d\dot{Z} \end{bmatrix} - [\partial r/\partial(r_o,C's,S's)]\begin{bmatrix} dr_o \\ dC's \\ dS's \end{bmatrix}$$

$$- [\partial r/\partial(X,Y,Z,\dot{X},\dot{Y},\dot{Z})]\begin{bmatrix} dX \\ dY \\ dZ \\ d\dot{X} \\ d\dot{Y} \\ d\dot{Z} \end{bmatrix} + [(R^0 - r^0 + d) - H^b].$$

Using (2.1) for the partial differentiation of R and (2.21) for the partial differentiation of (-r) and grouping these terms together, we obtain the following observation equation:

(2.22)

$$V = \{(X_s/R)[1-(a/R)e^2 \sin^2\bar{\phi}],\ (Y_s/R)[1-(a/R)e^2 \sin^2\bar{\phi}],\ (Z_s/R)[1+(a/R)e^2 \cos^2\bar{\phi}]\}$$

$$\times \tilde{M}_1 \begin{bmatrix} dX \\ dY \\ dZ \\ d\dot{X} \\ d\dot{Y} \\ d\dot{Z} \end{bmatrix} - [\partial r/\partial(r_o, C's, S's)] \begin{bmatrix} dr_o \\ dC's \\ dS's \end{bmatrix} + [(R^o - r^o + d) - H^b].$$

2.5 Summary and Conclusion

The observation equation corresponding to the elementary model was given in (2.4). This formula may be satisfactory for some preliminary reductions and for error propagation studies only. However, it is possible to extend its use and validity by including two kinds of corrections, one to be applied to the discrepancy term and the other, to the partial derivatives. To deduce these corrections, we made use of the ellipsoidal approximation to the geoid.

The correction to be added to the discrepancy term appeared in (2.8), namely

$$d \simeq (1/8)\, e^4\, r_0[H/(r_0+H)]\, \sin^2 2\theta.$$

Utilizing the numerical values of (2.6), (2.7), and (2.9), we may further approximate it as

$$d \simeq 4.9m \sin^2 2\theta \ \ldots \text{ per 1000 km of } H$$

or

$$d \simeq 4.9m \sin^2 2\bar{\phi} \ \ldots \text{ per 1000 km of } H,$$

where we have assumed that d is nearly linear in H for low satellites; for more accuracy, one could multiply such results by $7.37H/(r_0+H)$. The error in d resulting from the ellipsoidal approximation amounts to no more than a few centimeters in the worst case and it may be neglected (see the discussion that followed 2.5). For the reasons given earlier, the importance of the formula for d resides in its contribution to the proof of more rigorous partial derivatives seen below.

The second type of correction is to be applied to the directional cosines X_S/R, Y_S/R, Z_S/R appearing in the partial derivatives of R. These values should be multiplied by $[1-(a/R)\ e^2\ \sin^2\bar{\phi}]$, $[1-(a/R)\ e^2\ \sin^2\bar{\phi}]$, $[1+(a/R)\ e^2\ \cos^2\bar{\phi}]$, respectively. Taking $R \simeq a+H$, $a \simeq 6378$km, $H \simeq 1000$km, $e^2 \simeq 0.0067$, the second terms (corrections) in the brackets are respectively $-0.0058\ \sin^2\bar{\phi}$, $-0.0058\ \sin^2\bar{\phi}$, $+0.0058\ \cos^2\bar{\phi}$. Compared to unity, these corrections are small but not negligible. In particular, they may change the original partial derivatives by up to 0.6% of their values (on the average it is roughly 0.3%). If we did not apply them, the partial derivatives with respect to the state vector would often be good to only about 2 to 3 significant digits. By including them, we extend the validity of these partial derivatives perhaps by another 2 or more significant digits. The remaining errors stemming from the ellipsoidal approximation to the geoid are believed to be smaller than the above corrections by at least one order of magnitude and they diminish with decreasing deflections of the vertical. Furthermore, such remaining errors are random in nature unlike the corrections we have introduced. We may conclude that the accuracy of partial derivatives has been extended to at least four significant digits. This is compatible with the number of significant digits in the discrepancy terms, in view of the discussion that followed (2.10). Given the underlying assumption of observations over the oceanic regions referring to the geoid surface, it appears that the observation equations in the short arc mode of satellite altimetry formed according to (2.22) do not necessitate any further correction.

The practical usefulness of the refinement in the partial derivatives may be illustrated as follows. A small set of satellite altimetry data was adjusted by the AFGL program SARRA (Short Arc Reduction of Radar Altimetry). In this program, a different set of coefficients from C's and S's is used, but the state vector parameters are treated exactly as described earlier. After two iterations with the simplified partial derivatives, the parameters received no further corrections. However, using the refined partial derivatives, already the first adjustment yielded all the final values. Furthermore, it resulted in a very slightly different -- but more rigorous -- error propagation. It is conceivable that in this way iterations may be eliminated altogether, which will result in important savings in terms of computer run-time.

3. REALISTIC WEIGHTING OF THE STATE VECTOR PARAMETERS

3.1 General Considerations

One group of parameters in the short arc mode of satellite altimetry consists of the six state vector components (three position components and three velocity components) per short arc, associated with a point $\bar{S}$ approximately at mid-arc. The variance-covariance matrix of these six parameters is denoted as Σ when it is known in the earth fixed (E.F.) coordinate system of adjustment. The pertinent weight matrix of dimensions (6,6) is formed by inverting Σ. The weighting is independent from arc to arc.

It has been clearly established that the short arc mode of satellite altimetry does not require highly precise reference orbits. This being so, some earlier data reductions were made with unrealistically large a-priori standard errors (sigmas) of the state vector parameters, in the mode called "coarse error spheres"; the three position sigmas were equal and the three velocity sigmas were equal as well, and the matrix Σ was diagonal. The real error ellipsoids related to the state vectors were well within the boundaries of these spheres. However, it is well known that in reality, the error ellipsoids are far from having a spherical shape. In fact, they are elongated in the in-track component, the uncertainty being smaller in the cross-track component and the smallest in the third orthogonal component coinciding roughly with the "up" direction. The variance-covariance matrix in this system, denoted Σ_T, is diagonal. Brown [1975] reports that according to recent studies, the three positional sigmas of the Broadcast Ephemeris

may be characterized in this system by approximately 24m, 17m, and 8m, respectively. The three velocity sigmas used in some real data reductions at AFGL have been roughly estimated as 0.02m/sec, 0.02m/sec, and 0.01m/sec, respectively.

The passage from weights associated with the coarse error spheres to weights compatible with the realistic error ellipsoids has been one of the first tasks accomplished when developing and improving the satellite altimetry software. However, it has been implemented in two steps which have allowed some comparisons and verifications. In the first step, the six state vector sigmas have been assumed given in the "quasi-horizontal" (Q.H.) system associated with $\overline{S}$, the variance-covariance matrix in this system (Σ_{QH}), being diagonal. (The orthonormal base vectors in the Q. H. system are roughly parallel to those in the horizontal system associated with the foot-point of $\overline{S}$.) The matrix Σ has been obtained by the rigorous variance-covariance propagation from the Q.H. system to the E.F. system. The inversion of this matrix can be performed efficiently upon taking advantage of the fact that the transformation matrices are orthogonal. Exactly the same description can be applied to the second step, which differs from the above only in that the Q. H. system is replaced by the "track" system. These two steps will be described in two separate sections.

3.2 Sigmas of the State Vector Parameters Considered as Given in the Q.H. System

The components of the free vectors in the Q.H. system will be denoted as N, E, and U, corresponding roughly to the north, east, and up directions. This system is left-handed while the E.F. system is right-handed. We shall therefore first transform the components of vectors in the E.F. system into a right-handed X', Y', Z' system which corresponds to -N, E, U, and then simply change the signs associated with the first component. The relationship between the X, Y, Z (i.e., E.F.) system and the (X', Y', Z') system is shown in Figure 3.1, where the orthonormal base vectors associated with these two systems are denoted as $\vec{i}$, $\vec{j}$, $\vec{k}$ and $\vec{i}'$, $\vec{j}'$, $\vec{k}'$, respectively. The position vector of $\bar{S}$ is represented by $\vec{\rho}$.

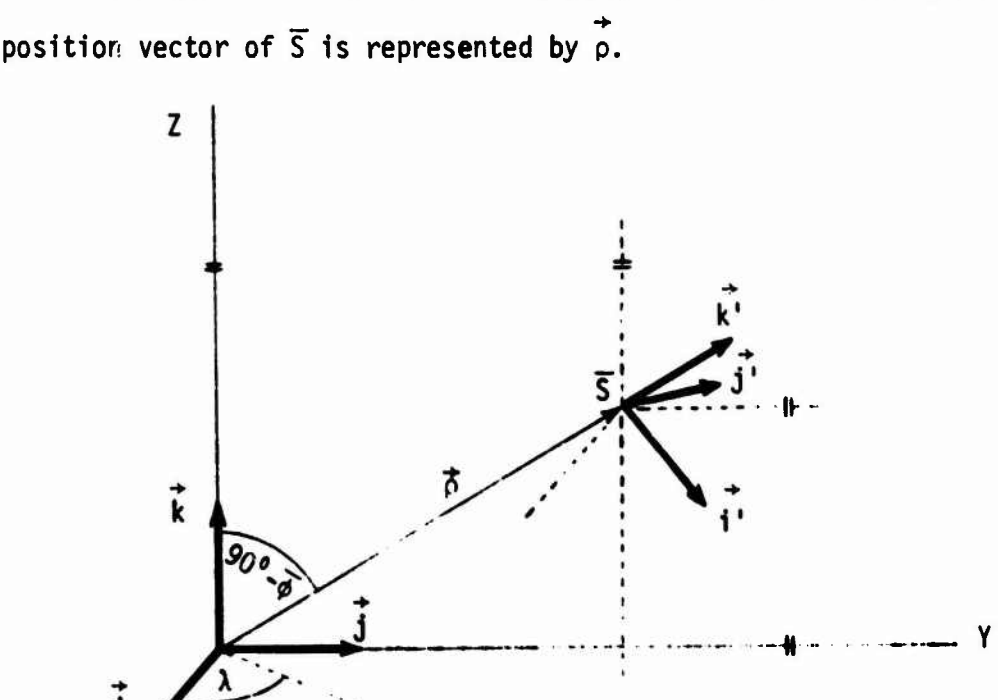


FIGURE 3.1

Relationship between the X, Y, Z and the X', Y', Z' systems

With the notations identifying the scalar products as

$$(i, j') \equiv \vec{i} \cdot \vec{j}' \quad , \quad \text{etc.} \quad ,$$

the transformation matrix between the E.F. and the X', Y', Z' systems may be written in general as follows:

$$\tilde{R}' = \begin{bmatrix} (i', i) & (i', j) & (i', k) \\ (j', i) & (j', j) & (j', k) \\ (k', i) & (k', j) & (k', k) \end{bmatrix} \quad , \tag{3.1}$$

so that (the free vectors are identified by Δ)

$$\begin{bmatrix} \Delta X' \\ \Delta Y' \\ \Delta Z' \end{bmatrix} = \tilde{R}' \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \quad . \tag{3.2}$$

The relationship (3.2) would apply also between the base vectors, i.e., if $\Delta X'$, $\Delta Y'$, $\Delta Z'$ were replaced by $\vec{i}'$, $\vec{j}'$, $\vec{k}'$, and if ΔX, ΔY, ΔZ were replaced by $\vec{i}$, $\vec{j}$, $\vec{k}$. When use is made of rotation matrices which in the present case are known, the formation of $\tilde{R}'$ is very simple. This procedure amounts essentially to decomposing the final transformation into two elementary transformations, and it is the subject of the next paragraph. However, we shall also proceed according to the general formula (3.1) because this is the route that will be followed when dealing with the "track" system.

The rotation matrices with respect to the second and third axes (in the E.F. system X, Y, and Z are identified with the axes 1, 2, and 3) are given as

$$R_2(\omega) = \begin{bmatrix} \cos\omega & 0 & -\sin\omega \\ 0 & 1 & 0 \\ \sin\omega & 0 & \cos\omega \end{bmatrix} ,$$

and

$$R_3(\omega) = \begin{bmatrix} \cos\omega & \sin\omega & 0 \\ -\sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix} ,$$

where the angle ω represents a positive rotation. If we want the orthonormal triad $\vec{i}, \vec{j}, \vec{k}$ to become parallel to its barred counterpart, it must first be rotated around the axis 3 by the angle λ, and this new system must then be rotated around its (new) axis 2 by the angle $90^0 - \bar{\phi}$, yielding

$$\tilde{R}' = R_2(90^0 - \bar{\phi})\ R_3(\lambda) .$$

Each of the three matrices above is orthogonal, i.e., its inverse equals its transpose as can be easily verified by direct computation. Carrying out the indicated multiplications, we obtain

$$\tilde{R}' = \begin{bmatrix} \sin\bar{\phi}\cos\lambda & \sin\bar{\phi}\sin\lambda & -\cos\bar{\phi} \\ -\sin\lambda & \cos\lambda & 0 \\ \cos\bar{\phi}\cos\lambda & \cos\bar{\phi}\sin\lambda & \sin\bar{\phi} \end{bmatrix} . \qquad (3.3)$$

The transformation matrix $\tilde{R}$ between the E.F. and the Q.H. systems would be formed from $\tilde{R}'$ by changing the signs of its first three elements (the first row). The transformation matrix between the E.F. and the customary horizontal system would be derived in precisely the same fashion except that $\bar{\phi}$ (the geocentric latitude) would be replaced by ϕ (the geodetic latitude).

In case we want to construct the transformation matrix directly from (3.1), we first generate the three base vectors sequentially as follows:

$$\begin{aligned} \vec{k}' &= \vec{\rho}/\rho \quad , \\ \vec{j}' &= \vec{k} \times \vec{k}'/c \quad , \\ \vec{i}' &= \vec{j}' \times \vec{k}' \quad , \end{aligned}$$

where "×" represents the vector product, and where

$$\rho = (X^2 + Y^2 + Z^2)^{\frac{1}{2}} \quad , \quad c = \rho_1/\rho \quad , \quad \rho_1 = (X^2 + Y^2)^{\frac{1}{2}} \quad .$$

With their components expressed in the E.F. system, the above base vectors are

$$\vec{k}' \equiv \begin{bmatrix} X/\rho \\ Y/\rho \\ Z/\rho \end{bmatrix} \quad , \quad \vec{j}' \equiv \begin{bmatrix} -Y/\rho_1 \\ X/\rho_1 \\ 0 \end{bmatrix} \quad , \quad \vec{i}' \equiv \begin{bmatrix} XZ/\rho\rho_1 \\ YZ/\rho\rho_1 \\ -\rho_1/\rho \end{bmatrix} ,$$

where

$$\vec{\rho} \equiv \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \qquad (3.4a)$$

has been utilized. Taking advantage of the fact that in the E.F. system, the vectors $\vec{i}, \vec{j}, \vec{k}$ are symbolized by

$$\vec{i} \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{j} \equiv \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{k} \equiv \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \qquad (3.4b)$$

we can immediately write the transformation matrix (3.1) as

$$\tilde{R}' = \begin{bmatrix} (Z/\rho)(X/\rho_1) & (Z/\rho)(Y/\rho_1) & -\rho_1/\rho \\ -Y/\rho_1 & X/\rho_1 & 0 \\ X/\rho & Y/\rho & Z/\rho \end{bmatrix} .$$

Upon realizing that

$$X/\rho = \cos\bar{\phi}\cos\lambda, \quad Y/\rho = \cos\bar{\phi}\sin\lambda, \quad Z/\rho = \sin\bar{\phi},$$
$$X/\rho_1 = \cos\lambda, \quad Y/\rho_1 = \sin\lambda, \quad \rho_1/\rho = \cos\bar{\phi},$$

we may verify equation (3.3).

We now change the signs associated with $\Delta X'$ and the first row of $\tilde{R}'$ in (3.2), yielding

$$\tilde{R} = \begin{bmatrix} -\sin\bar{\phi}\cos\lambda & -\sin\bar{\phi}\sin\lambda & \cos\bar{\phi} \\ -\sin\lambda & \cos\lambda & 0 \\ \cos\bar{\phi}\cos\lambda & \cos\bar{\phi}\sin\lambda & \sin\bar{\phi} \end{bmatrix}, \qquad (3.5)$$

and

$$\begin{bmatrix} N \\ E \\ U \end{bmatrix} = \tilde{R} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \quad . \tag{3.6}$$

Equation (3.6), although an auxiliary feature in the present derivation, is utilized in the program SAGG to express the first three state vector corrections in the Q.H. system; these positional corrections are computed as ΔX, ΔY, ΔZ in the E.F. system. Exactly the same statement can be made of course with regard to the three velocity corrections identified by dots, namely

$$\begin{bmatrix} \dot{N} \\ \dot{E} \\ \dot{U} \end{bmatrix} = \tilde{R} \begin{bmatrix} \Delta\dot{X} \\ \Delta\dot{Y} \\ \Delta\dot{Z} \end{bmatrix} \quad . \tag{3.7}$$

Since for the full correction vector we have

$$\begin{bmatrix} N \\ E \\ U \\ \dot{N} \\ \dot{E} \\ \dot{U} \end{bmatrix} = \begin{bmatrix} \tilde{R} & 0 \\ 0 & \tilde{R} \end{bmatrix} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \\ \Delta\dot{X} \\ \Delta\dot{Y} \\ \Delta\dot{Z} \end{bmatrix} \quad , \tag{3.8}$$

the a-posteriori variance-covariance matrix in the Q.H. system is

$$\Sigma_{QH} = \begin{bmatrix} \tilde{R} & 0 \\ 0 & \tilde{R} \end{bmatrix} \Sigma \begin{bmatrix} \tilde{R} & 0 \\ 0 & \tilde{R} \end{bmatrix}^T , \qquad (3.9)$$

where "T" implies that transposition and where the matrix on the left-hand side as well as all three matrices on the right-hand side are of dimensions (6,6). The square roots of the diagonal elements in Σ_{QH} are the a-posteriori sigmas of the state vector parameters in the Q.H. system (they are also printed by SAGG).

However, we are not concerned with the a-posteriori sigmas of the state vector parameters. Instead, the topic of this section has been to derive their weight matrix (P_x) in the E.F. system (i.e., in the system of adjustment). The matrix P_x is obtained by inverting the matrix Σ, containing this time the a-priori variances-covariances; Σ is not given, but must be computed from the a-priori variance-covariance matrix Σ_{QH}. Most often, this matrix Σ_{QH} is diagnal, consisting of the squares of individual a-priori sigmas as presented symbolically below:

$$\Sigma_{QH} = \text{diag}\,(\sigma_N^2, \sigma_E^2, \sigma_U^2; \sigma_{\dot{N}}^2, \sigma_{\dot{E}}^2, \sigma_{\dot{U}}^2) , \qquad (3.10)$$

or

$$\Sigma_{QH} = \text{diag}\,(d_{QH}, \dot{d}_{QH}) , \qquad (3.11a)$$

where

$$d_{QH} = \text{diag}\,(\sigma_N^2, \sigma_E^2, \sigma_U^2) , \qquad (3.11b)$$

$$\dot{d}_{QH} = \text{diag}\,(\sigma_{\dot{N}}^2, \sigma_{\dot{E}}^2, \sigma_{\dot{U}}^2) . \qquad (3.11c)$$

Since the matrix $\tilde{R}$ is orthogonal, from (3.9) we deduce that

$$\Sigma = \begin{bmatrix} \tilde{R}^T & 0 \\ 0 & \tilde{R}^T \end{bmatrix} \Sigma_{QH} \begin{bmatrix} \tilde{R} & 0 \\ 0 & \tilde{R} \end{bmatrix} . \qquad (3.12a)$$

In view of (3.10) or (3.11), this can be simplified as

$$\Sigma = \begin{bmatrix} \tilde{R}^T d_{QH} \tilde{R} & 0 \\ 0 & \tilde{R}^T \dot{d}_{QH} \tilde{R} \end{bmatrix} . \qquad (3.12b)$$

We observe that although Σ is not diagonal, it contains two diagonal blocks of dimensions (3,3). The three position components remain therefore independent from the three velocity components also in the E.F. system. As a minor remark, we notice that from the properties of the trace operator (Tr), it follows e.g. from (3.12a) that

$$Tr \left(\Sigma\right) = Tr \left(\Sigma_{QH}\right) ;$$

a similar equality holds also with respect to either of the two diagonal blocks. Thus, we have for example

$$\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 = \sigma_N^2 + \sigma_E^2 + \sigma_U^2 .$$

Due to the quasi-diagonal form of (3.12b) where $\tilde{R}$ is an orthogonal matrix, the weight matrix can be conputed efficiently as

$$P_x \equiv \Sigma^{-1} = \begin{bmatrix} \tilde{R}^T d_{QH}^{-1} \tilde{R} & 0 \\ 0 & \tilde{R}^T \dot{d}_{QH}^{-1} \tilde{R} \end{bmatrix} , \qquad (3.13)$$

where, in analogy to (3.11),

$$d_{QH}^{-1} = \text{diag}\,(\sigma_N^{-2},\ \sigma_E^{-2},\ \sigma_U^{-2}) \quad , \tag{3.14a}$$

$$\dot{d}_{QH}^{-1} = \text{diag}\,(\sigma_{\dot{N}}^{-2},\ \sigma_{\dot{E}}^{-2},\ \sigma_{\dot{U}}^{-2}) \quad . \tag{3.14b}$$

Since hundreds of short arcs may be processed in a single adjustment, it is imperative to perform the pertinent inversions analytically as in (3.13), in case the a-priori sigmas are indeed given in the Q.H. system. If the orbit is circular, this system is equivalent to the "track" system to be described next.

3.3 Sigmas of the State Vector Parameters Given in the "Track" System

The first two out of the three base vectors in the "track" (T) system coincide with the in-track and the cross-track directions, respectively. The third base vector forms a left-handed orthonormal triad with the first two; if the orbit is not too elongated, it corresponds roughly to the "up" direction. The components of free vectors in this system will be denoted by the capital letters I (in-track), C (cross-track), and O (orthogonal to I, C). However, to relate the T system to the E.F. system, we shall use a right-handed intermediate system X", Y", Z", much in the same way as the system X', Y', Z', has been used in the previous section. This intermediate system corresponds to -I, C, O; thus the desired expressions will be again obtained by simply changing the signs associated with the first component. Figure 3.2 relates the X", Y", Z" system represented by its base vectors $\vec{i}$", $\vec{j}$", $\vec{k}$" to the X, Y, Z (i.e., E.F.) system, depicted also in Figure 3.1. As an additional feature, Figure 3.2 displays the velocity vector $\vec{v}$. The components of $\vec{\rho}$ and $\vec{v}$ are given in the E.F. system as

$$\vec{\rho} \equiv \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} , \qquad \vec{v} \equiv \begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} ; \qquad (3.15)$$

the first of these two relations has been utilized previously (see equation 3.4a). The two vectors in (3.15) form the state vector.

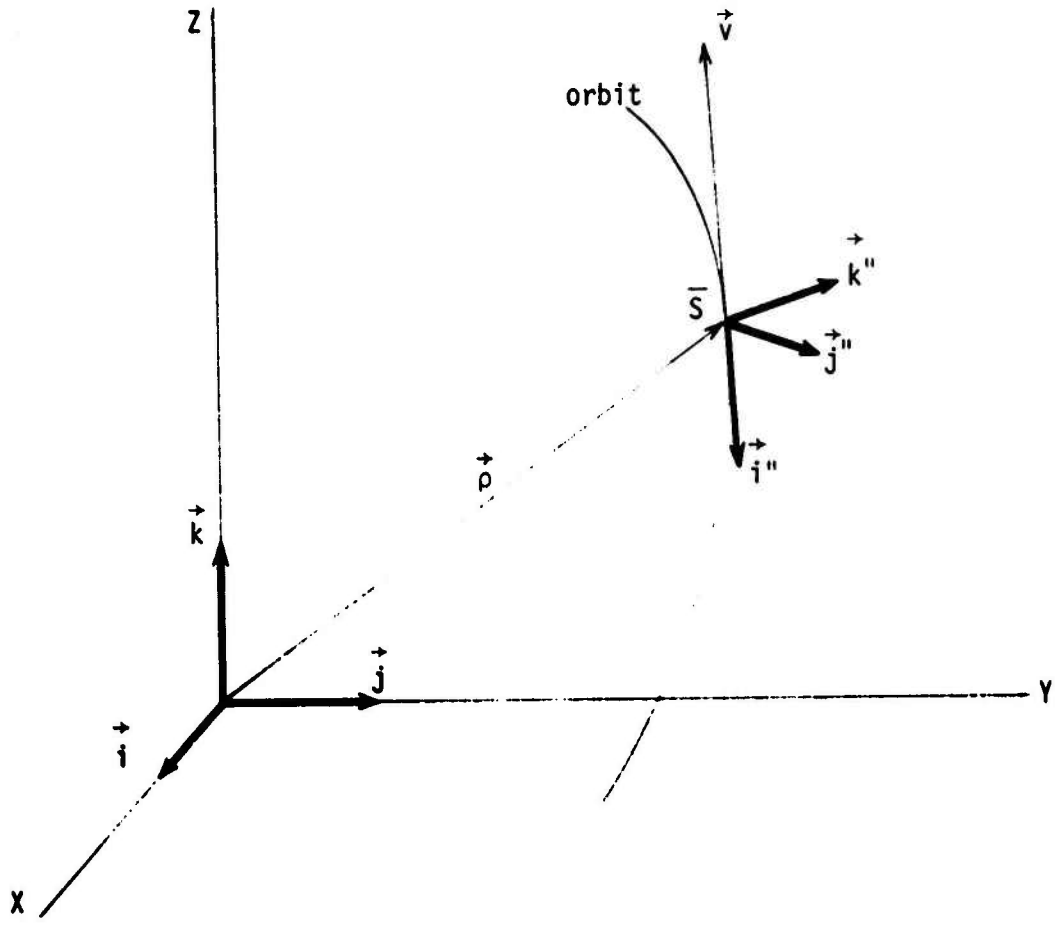


FIGURE 3.2

Relationship between the X, Y, Z, and the X", Y", Z" systems

In the present derivation, only the direct approach corresponding to (3.1) will be utilized. In analogy to (3.1) and (3.2), we now have

$$R'' = \begin{bmatrix} (i'', i) & (i'', j) & (i'', k) \\ (j'', i) & (j'', j) & (j'', k) \\ (k'', i) & (k'', j) & (k'', k) \end{bmatrix}, \qquad (3.16)$$

$$\begin{bmatrix} \Delta X'' \\ \Delta Y'' \\ \Delta Z'' \end{bmatrix} = R'' \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \quad . \tag{3.17}$$

The three base vectors are expressed sequentially as

$$\begin{aligned} \vec{i}'' &= -\vec{v}/v \quad , \\ \vec{j}'' &= \vec{v} \times \vec{\rho}/h \quad , \\ \vec{k}'' &= \vec{i}'' \times \vec{j}'' \quad , \end{aligned}$$

where

$$v = (\dot{X}^2 + \dot{Y}^2 + \dot{Z}^2)^{\frac{1}{2}} \quad , \tag{3.18}$$

and where h is the scale factor (to be expressed later) that makes $\vec{j}''$ a unit vector.

The components of the above base vectors can be expressed in the E.F. system in a straightforward fashion:

$$\vec{i}'' \equiv \begin{bmatrix} -\dot{X}/v \\ -\dot{Y}/v \\ -\dot{Z}/v \end{bmatrix} \quad ;$$

$$\vec{j}'' \equiv \begin{bmatrix} A_1/h \\ A_2/h \\ A_3/h \end{bmatrix} \quad ,$$

where

$$\left.\begin{aligned} A_1 &= \dot{Y}Z - \dot{Z}Y \ , \\ A_2 &= \dot{Z}X - \dot{X}Z \ , \\ A_3 &= \dot{X}Y - \dot{Y}X \ , \end{aligned}\right\} \qquad (3.19a)$$

and where

$$h = (A_1^2 + A_2^2 + A_3^2)^{\frac{1}{2}} \ ; \qquad (3.19b)$$

$$\vec{k}'' = \begin{bmatrix} B_1 / q \\ B_2 / q \\ B_3 / q \end{bmatrix} \ ,$$

where

$$\left.\begin{aligned} B_1 &= A_2 \dot{Z} - A_3 \dot{Y} \ , \\ B_2 &= A_3 \dot{X} - A_1 \dot{Z} \ , \\ B_3 &= A_1 \dot{Y} - A_2 \dot{X} \ , \end{aligned}\right\} \qquad (3.20a)$$

and where

$$q = vh \ . \qquad (3.20b)$$

Alternatively, from (3.19a) and (3.20a) we may also express

$$\begin{aligned} B_1 &= X(\dot{Y}^2 + \dot{Z}^2) - \dot{X}(Y\dot{Y} + Z\dot{Z}) \ , \\ B_2 &= Y(\dot{Z}^2 + \dot{X}^2) - \dot{Y}(Z\dot{Z} + X\dot{X}) \ , \\ B_3 &= Z(\dot{X}^2 + \dot{Y}^2) - \dot{Z}(X\dot{X} + Y\dot{Y}) \ . \end{aligned}$$

It can be verified that

$$q = (B_1^2 + B_2^2 + B_3^2)^{\frac{1}{2}}$$

as it should. Considering (3.4b), from (3.16), we find

$$R'' = \begin{bmatrix} -\dot{X} / v & -\dot{Y} / v & -\dot{Z} / v \\ A_1 / h & A_2 / h & A_3 / h \\ B_1 / q & B_2 / q & B_3 / q \end{bmatrix} .$$

Upon changing the signs associated with $\Delta X''$ and the first row of R'' in (3.17), one obtains

$$R = \begin{bmatrix} \dot{X} / v & \dot{Y} / v & \dot{Z} / v \\ A_1 / h & A_2 / h & A_3 / h \\ B_1 / q & B_2 / q & B_3 / q \end{bmatrix} , \qquad (3.21)$$

and

$$\begin{bmatrix} I \\ C \\ O \end{bmatrix} = R \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} . \qquad (3.22a)$$

The same transformation holds true for the velocity components, namely

$$\begin{bmatrix} \dot{I} \\ \dot{C} \\ \dot{O} \end{bmatrix} = R \begin{bmatrix} \Delta\dot{X} \\ \Delta\dot{Y} \\ \Delta\dot{Z} \end{bmatrix} . \qquad (3.22b)$$

As a matter of verification, one can show by straightforward algebra that

$$R\,R^T = I$$

is indeed fulfilled as it should be for orthogonal matrices. Due to this property, we can also express

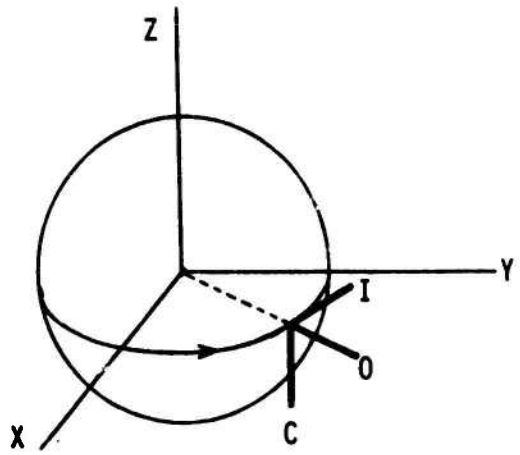
$$\begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} = R^T \begin{bmatrix} I \\ C \\ 0 \end{bmatrix} , \tag{3.22c}$$

$$\begin{bmatrix} \dot{\Delta X} \\ \dot{\Delta Y} \\ \dot{\Delta Z} \end{bmatrix} = R^T \begin{bmatrix} \dot{I} \\ \dot{C} \\ \dot{0} \end{bmatrix} . \tag{3.22d}$$

A simple example that to a certain extent verifies the formulas for the Q.H. system and the T system may be constructed in the following manner. Suppose that a circular orbit on the equator should be treated in both systems. From the symbolic sketch below, one can see immediately that the result

$$\begin{bmatrix} I \\ C \\ 0 \end{bmatrix} = \begin{bmatrix} E \\ -N \\ U \end{bmatrix}$$

should be obtained. The angular velocity is given by $k > 0$.



The state vector components (in the E.F. system) are

$$(X,\ Y,\ Z;\ \dot{X},\ \dot{Y},\ \dot{Z}) \equiv (X,\ Y,\ 0;\ -kY,\ +kX,\ 0) \quad .$$

The velocity is expressed as

$$v = k\,(X^2 + Y^2)^{\frac{1}{2}} \quad ;$$

according to (3.19) and (3.20), we obtain

$$A_1 = 0 \quad ,$$
$$A_2 = 0 \quad ,$$
$$A_3 = -k\,(X^2 + Y^2) \quad ;$$

$$h = k\,(X^2 + Y^2) \quad ;$$

$$B_1 = k^2 X\,(X^2 + Y^2) \quad ,$$
$$B_2 = k^2 Y\,(X^2 + Y^2) \quad ,$$
$$B_3 = 0 \quad ;$$

$$q = k^2 \, (X^2 + Y^2)^{3/2} \quad .$$

From (3.21) and (3.22a), we then have

$$R = \begin{bmatrix} -Y / (X^2 + Y^2)^{\frac{1}{2}} & X / (X^2 + Y^2)^{\frac{1}{2}} & 0 \\ 0 & 0 & -1 \\ X / (X^2 + Y^2)^{\frac{1}{2}} & Y / (X^2 + Y^2)^{\frac{1}{2}} & 0 \end{bmatrix} ,$$

$$\begin{bmatrix} I \\ C \\ O \end{bmatrix} = R \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \quad .$$

As for the horizontal system, this example is characterized by

$$\sin\bar{\phi} = 0 \quad , \qquad \cos\bar{\phi} = 1 \quad ,$$

$$\sin\lambda = Y / (X^2 + Y^2)^{\frac{1}{2}}, \quad \cos\lambda = X / (X^2 + Y^2)^{\frac{1}{2}} \; ;$$

thus (see 3.5, 3.6)

$$\tilde{R} = \begin{bmatrix} 0 & 0 & 1 \\ -Y / (X^2 + Y^2)^{\frac{1}{2}} & X / (X^2 + Y^2)^{\frac{1}{2}} & 0 \\ X / (X^2 + Y^2)^{\frac{1}{2}} & Y / (X^2 + Y^2)^{\frac{1}{2}} & 0 \end{bmatrix} ,$$

$$\begin{bmatrix} N \\ E \\ U \end{bmatrix} = \tilde{R} \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \quad .$$

If we change the sign of N and then exchange it with E, the corresponding changes in $\tilde{R}$ transform it into R which gave rise to the components I, C, O. In particular,

$$\begin{bmatrix} E \\ -N \\ U \end{bmatrix} = R \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \equiv \begin{bmatrix} I \\ C \\ O \end{bmatrix} ,$$

which verifies the desired result and concludes our example.

In order to carry out the variance-covariance propagation, we write the relations (3.22c) and (3.22d) in a compact form as

$$\begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \\ \Delta \dot{X} \\ \Delta \dot{Y} \\ \Delta \dot{Z} \end{bmatrix} = \begin{bmatrix} R^T & 0 \\ 0 & R^T \end{bmatrix} \begin{bmatrix} I \\ C \\ O \\ \dot{I} \\ \dot{C} \\ \dot{O} \end{bmatrix} ,$$

which yields

$$\Sigma = \begin{bmatrix} R^T & 0 \\ 0 & R^T \end{bmatrix} \Sigma_T \begin{bmatrix} R & 0 \\ 0 & R \end{bmatrix} . \qquad (3.23)$$

The "track" system is of course the one in which the realistic a-priori sigmas of the state vector components are available. The matrix Σ_T is usually

diagonal; in analogy to (3.11), we have

$$\Sigma_T = \text{diag}\,(d_T, \dot{d}_T) \quad , \tag{3.24a}$$

where

$$d_T = \text{diag}\,(\sigma_I^2, \sigma_C^2, \sigma_O^2) \quad , \tag{3.24b}$$

$$\dot{d}_T = \text{diag}\,(\sigma_{\dot{I}}^2, \sigma_{\dot{C}}^2, \sigma_{\dot{O}}^2) \quad . \tag{3.24c}$$

The relation (3.23) can thus be simplified as follows:

$$\Sigma = \begin{bmatrix} R^T\, d_T\, R & 0 \\ 0 & R^T\, \dot{d}_T\, R \end{bmatrix} \quad . \tag{3.25}$$

The a-priori matrix Σ contains again two diagonal blocks; the three position components remain therefore independent from the three velocity components also in the E.F. system. One can again observe that

$$\text{Tr}\,(\Sigma) = \text{Tr}\,(\Sigma_T) \quad ,$$

$$\sigma_X^2 + \sigma_Y^2 + \sigma_Z^2 = \sigma_I^2 + \sigma_C^2 + \sigma_O^2 \text{ , etc.}$$

In a fashion completely analogous to (3.13), we form the weight matrix efficiently as

$$P_X \equiv \Sigma^{-1} = \begin{bmatrix} R^T\, d_T^{-1}\, R & 0 \\ 0 & R^T\, \dot{d}_T^{-1}\, R \end{bmatrix} \quad , \tag{3.26}$$

where

$$d_T^{-1} = \text{diag}\,(\sigma_I^{-2}, \sigma_C^{-2}, \sigma_O^{-2}) \quad , \tag{3.27a}$$

$$\dot{d}_T^{-1} = \text{diag}\,(\sigma_{\dot{I}}^{-2}, \sigma_{\dot{C}}^{-2}, \sigma_{\dot{O}}^{-2}) \quad . \tag{3.27b}$$

The matrix R has been presented in (3.21) with v, h, and q expressed explicitly in (3.18), (3.19b) and (3.20b), respectively, with A_1, A_2, A_3 given in (3.19a), and with B_1, B_2, B_3 given in (3.20a). If the a-priori matrix Σ could ever become directly available, the weight matrix would be formed by simply inverting it (see the first part of 3.26). However, the best procedure in practice consists in acquiring a realistic a-priori information, for a given satellite, about the uncertainty in the state vector components. Such information has the form of six independent sigmas in the "track" system. It can then be utilized according to equation (3.26), which is also the procedure implemented in the program SAGG.

4. PRACTICAL CONSIDERATIONS RELATED TO SATELLITE ALTIMETRY

4.1 Observation Equations in the Short Arc Satellite Altimetry Adjustment Model

In this section we shall present the observation equations of Chapter 2 in more detail. In agreement with Chapter 1, r_0 will be a fixed quantity rather than an adjustable parameter. Our main task will consist in expressing explicitly the partial derivatives $\partial r/\partial(C's, S's)$ that figured in equation (2.22). In addition, more information leading to a practical computation of the angle θ and of the matrix $\tilde{M}_1$ will be presented.

In agreement with our earlier description, the state vector components of an arc are symbolized by X, Y, Z, $\dot{X}$, $\dot{Y}$, $\dot{Z}$, and the coordinates of S, a point of this arc, by X_S, Y_S, Z_S (all in the E.F. system). The epoch (of the state vector) is denoted by t_0 and the time of the "event S" is t_S. The position of S for a given $t_S - t_0$ is a function of the state vector, written as

$$X_S = X_S(X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}),$$
$$Y_S = Y_S(X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}),$$
$$Z_S = Z_S(X, Y, Z, \dot{X}, \dot{Y}, \dot{Z}).$$

We have seen that this position is considered to be independent of the adjustable potential coefficients C's, and S's. From the coordinates of S, the two geocentric coordinates θ, λ of the geoidal point P (a sub-satellite point) are obtained as follows. First, the geodetic latitude may be

found through an iterative process as

$$\text{tg}\phi = \left| Z_S/(X_S^2 + Y_S^2)^{\frac{1}{2}} \right| \; (N + h) / \left| N(1 - e^2) + h \right| ,$$

$$N = a / (1 - e^2 \sin^2\phi)^{\frac{1}{2}} ,$$

$$h = (X_S^2 + Y_S^2)^{\frac{1}{2}}/\cos\phi - N ,$$

and the longitude, from the simple formula

$$\text{tg}\lambda = Y_S / X_S .$$

The position of P is thus computed with the aid of the measured satellite altitude by the three equations that followed (2.9"). The geocentric angle θ is then obtained from (2.9*).

The angles θ, λ having been computed, we may proceed to express r and the pertinent partial derivatives. According to Figure 2.2 or the statement that followed equation (2.9*), r is given by (2.2) upon replacing $\bar{\phi}$ by θ. We have

$$r = r_o \left[1 + \sum_{n=2}^{\infty} (a/r)^n \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P_{nm}(\sin\theta) \right] + \tfrac{1}{2} \omega^2 r_o r^3 \cos^2\theta/(kM) , \qquad (4.1)$$

which yields

$$dr = -(r_o/r) \sum_{n=2}^{\infty} n(a/r)^n \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P_{nm}(\sin\theta) dr + r_c \sum_{n=2}^{\infty} (a/r)^n \sum_{m=0}^{n} (dC_{nm} \cos m\lambda + dS_{nm} \sin m\lambda) P_{nm}(\sin\theta) + (3/2)(r_o/r)\omega^2 r \left[r^2/(kM) \right] \cos^2\theta \, dr . \qquad (4.2)$$

We next introduce the following notations:

$$S(n) = \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P_{nm}(\sin\theta), \tag{4.3a}$$

$$P_1 = \sum_{n=2}^{\infty} (a/r)^n S(n), \tag{4.3b}$$

$$P_2 = \sum_{n=2}^{\infty} n(a/r)^n S(n), \tag{4.3c}$$

$$C = kM / r^2, \tag{4.3d}$$

$$D = \omega^2 r (1/C) \cos^2\theta . \tag{4.3e}$$

The formula (4.1) is thus written as

$$r = r_o (1 + P_1 + \tfrac{1}{2}D) ; \tag{4.4}$$

on the other hand, the formula (4.2) yields

$$Qdr = r_o \sum_{n=2}^{\infty} (a/r)^n \sum_{m=0}^{n} (dC_{nm} \cos m\lambda + dS_{nm} \sin m\lambda) \times P_{nm}(\sin\theta) , \tag{4.5a}$$

where

$$Q = 1 + (r_o / r)(P_2 - 3D/2) . \tag{4.5b}$$

With the aid of (4.5), the required partial derivatives are readily formed as follows ($m > 0$ is assumed):

$$\partial r/\partial C_{n0} = r_o (a/r)^n P_n (\sin\theta) / Q , \tag{4.6a}$$

$$\partial r/\partial C_{nm} = r_o (a/r)^n P_{nm}(\sin\theta) \cos m\lambda / Q , \tag{4.6b}$$

$$\partial r/\partial S_{nm} = r_o (a/r)^n P_{nm}(\sin\theta) \sin m\lambda / Q . \tag{4.6c}$$

The observation equations (2.22) can now be rewritten more explicitly. In particular, we have

$$V = \{(X_s/R)[1-(a/R)e^2\sin^2\bar{\phi}],\ (Y_s/R)[1-(a/R)e^2\sin^2\bar{\phi}],(Z_s/R)[1$$

$$+(a/R)e^2\cos^2\bar{\phi}]\} \times \tilde{M}_1 \begin{bmatrix} dX \\ dY \\ dZ \\ d\dot{X} \\ d\dot{Y} \\ d\dot{Z} \end{bmatrix} - \left[\ldots \partial r/\partial C_{n0}\ldots,\ \ldots \partial r/\partial C_{nm}\ldots,\right.$$

$$\left.\ldots \partial r/\partial S_{nm}\ldots\right] \begin{bmatrix} \vdots \\ dC_{n0} \\ \vdots \\ dC_{nm} \\ \vdots \\ dS_{nm} \\ \vdots \end{bmatrix} + L \quad , \qquad (4.7)$$

where the partial derivatives of r with respect to the potential coefficients have been given in (4.6). The truncation in all of the above formulas are made at a suitable degree and order n = N. The "discrepancy" term L can be expressed as

$$L = OP^0 - r^0 \quad , \qquad (4.8)$$

where OP^0 is computed according to (2.9') or (2.9") via the approximate state vector parameters and the measured altimetry value , and where r^0 is computed from the initial values of the potential coefficients. We mention that the form of L in (2.22) corresponded to (2.9').

The matrix $\tilde{M}_1$ appearing in (4.7) has been described prior to (2.1). The first and the last of the three matrices forming $\tilde{M}_1$ will now be given explicitly. In particular, upon using an advantageous choice of the inertial system for each arch so that it coincides with the E.F. system at the epoch (t_0), one can express

$$\frac{\partial (X_s, Y_s, Z_s)}{\partial (x_s, y_s, z_s)} = \begin{bmatrix} \cos [\omega(t_s - t_0)] & \sin [\omega(t_s - t_0)] & 0 \\ -\sin [\omega(t_s - t_0)] & \cos [\omega(t_s - t_0)] & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.9a)$$

and

$$\frac{\partial (x, y, z, \dot{x}, \dot{y}, \dot{z})}{\partial (X, Y, Z, \dot{X}, \dot{Y}, \dot{Z})} = \begin{bmatrix} I & 0 \\ T & I \end{bmatrix}, \quad T = \begin{bmatrix} 0 & -\omega & 0 \\ \omega & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (4.9b)$$

This concludes the detailed presentation of the observation equations. One should keep in mind that a given set of state vector parameters remains the same only for the events on one arc, whereas the set of potential coefficients does not change from arc to arc.

4.2 Computation of Geoid Undulations from Radial Distances to the Geoid

We shall derive an efficient, yet accurate formula for the separation between the reference ellipsoid and P, a point on the geoid. The pertinent situation is depicted in Figure 4.1, where the rigorous geoid undulation is denoted by N. We shall prove that for all practical purposes, the distance $N \equiv PQ$ can be computed as the distance $N' \equiv PQ'$, and shall give the means to express the latter. In order to obtain the desired results, we shall specialize a few formulas from Appendix 1 to the present situation.

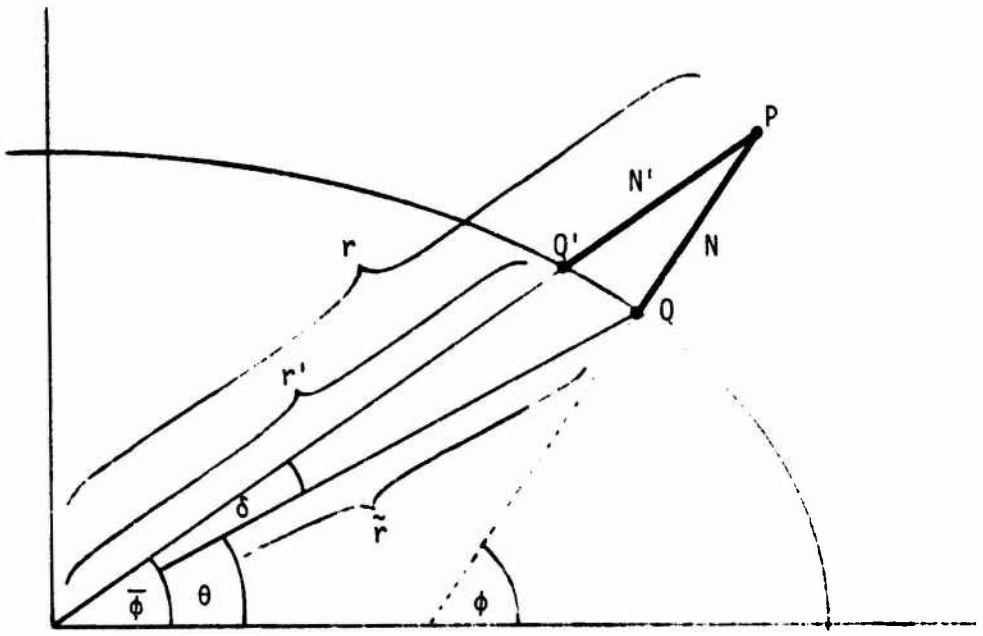


FIGURE 4.1

Meridian section of the reference ellipsoid showing the geoid undulation N

The basic formula for N follows from specializing (A1.16):

$$N = r - \tilde{r} + d \quad , \tag{4.10}$$

where all the quantities except d have been shown in Figure 4.1. Upon specializing (A1.20), one finds

$$d \simeq \tfrac{1}{2} \delta^2 R^2 / N \quad , \tag{4.11}$$

where R (the earth's mean radius) can be adopted as 6371 km. Finally, if we also specialize (A1.14) or (A1.14'), we can write to a good approximation:

$$\delta \simeq \tfrac{1}{2} e^2 (N / R) \sin 2\theta \quad ,$$

or, to about the same degree of accuracy,

$$\delta \simeq \tfrac{1}{2} e^2 (N / R) \sin 2\phi \quad ; \tag{4.12}$$

for our task, the value $e^2 \simeq 0.006695$ is sufficiently accurate. If (4.12) is squared and substituted into (4.11), if follows that

$$d \simeq (1/8) e^4 N \sin^2 2\phi \quad . \tag{4.13}$$

We now rewrite (4.10) as

$$N = r' + N' - \tilde{r} + d \quad ,$$

or

$$N = N' + \Delta \quad , \tag{4.14}$$

where

$$\Delta = dr + d \quad , \tag{4.15}$$

with

$$dr \equiv r' - \tilde{r} = (\partial r / \partial \phi)\, d\phi \quad .$$

The partial derivative in this last relation has been given in (2.15'); in particular, we use

$$\partial r / \partial \phi \simeq - \tfrac{1}{2} Re^2 \sin 2\phi \quad .$$

The value of $d\phi$ is given with sufficient accuracy by (4.12). We thus deduce that

$$dr \simeq - (1/4)\, e^4 N \sin^2 2\phi \quad .$$

But comparing this expression with (4.13), we observe that

$$dr \simeq - 2d \quad ,$$

and thus, from (4.15), that

$$\Delta \simeq - (1/8)\, e^4 N \sin^2 2\phi \quad . \qquad (4.16)$$

However, even if we took N as large as 120 m, the value of Δ would be negligible. In particular, we would have

$$\Delta \simeq - 0.672 \text{ mm } \sin^2 2\phi \quad .$$

Thus, for all practical purposes we can write

$$N = N' \quad , \qquad (4.17)$$

where

$$N' = r - r' \quad . \qquad (4.18)$$

For given geocentric coordinates $\overline{\phi}$, λ, the quantity r can be computed from the potential coefficients as in equation (2.2). On the other hand, a simple formula for the ellipsoidal radial distance r' in terms of $\overline{\phi}$ can be derived in a straightforward manner, as is shown below.

We begin with equation (2.15), which in our present notation is

$$r' = a(1 - e^2 \sin^2\phi)^{-\frac{1}{2}} [1 - e^2 (2 - e^2) \sin^2 \phi]^{\frac{1}{2}} \quad . \qquad (4.19)$$

To develop r' further, we need a few identities, such as

$$\sin^2\overline{\phi} = \mathrm{tg}^2\overline{\phi} / (1 + \mathrm{tg}^2 \overline{\phi})$$

and (see e.g. A1.6; ϕ in 4.19 or below is not the same as in Figure 4.1)

$$\mathrm{tg}\overline{\phi} = (1 - e^2) \mathrm{tg}\, \phi \quad ;$$

squaring the latter equation and substituting it into the former, we obtain

$$\sin^2\overline{\phi} = (1 - e^2)^2 \mathrm{tg}^2\phi / [1 + (1 - e^2)^2 \mathrm{tg}^2\phi] \quad . \qquad (4.20)$$

Further we have

$$1 - e^2 (2 - e^2) \sin^2\phi = \cos^2\phi + \sin^2\phi - 2e^2 \sin^2\phi + e^4 \sin^2 \phi$$
$$= \cos^2\phi [1 + (1 - e^2)^2 \mathrm{tg}^2\phi] \quad , \qquad (4.21)$$

and

$$1 - e^2 \sin^2\phi \equiv 1 - 2e^2 \sin^2\phi + e^4 \sin^2\phi + e^2 \sin^2\phi - e^4 \sin^2\phi$$
$$= 1 - e^2(2 - e^2) \sin^2\phi + e^2(1 - e^2) \sin^2\phi \; . \qquad (4.22)$$

Upon combining $1/r'^2$ given by (4.19) with (4.22), we find that

$$1/r'^2 = (1/a^2)\,\{1 + e^2\,(1 - e^2)\,\sin^2\phi\,/\,[1 - e^2(2 - e^2)\,\sin^2\phi]\}.$$

If (4.21) is now introduced into this equation, one obtains

$$1/r'^2 = (1/a^2)\,\{1 + e^2(1 - e^2)\,tg^2\phi/[1 + (1 - e^2)^2\,tg^2\phi\,]\}\,.$$

Finally, considering (4.20), we have

$$1/r'^2 = (1/a^2)\,[1 + e^2\,\sin^2\bar{\phi}\,/(1 - e^2)\,]\,,$$

or

$$r' = a/\,[1 + e^2\,\sin^2\bar{\phi}\,/(1 - e^2)\,]^{\frac{1}{2}}\,. \qquad (4.23)$$

4.3 Improvements and Verifications of the Orbital Part of the Program SAGG

Satellite altimetry observations in SAGG have functional relationships with two kinds of parameters. In the first group of parameters are found the spherical harmonic potential coefficients (C's and S's) and in the second, the state vector components for each short arc (orbital parameters). The part of SAGG dealing with the recovery of C's and S's has been thoroughly tested through the method of "parameter shifts", described e.g. in Section 5.2, item "f", of [Blaha, 1975]. This has been done in case of a ficticious nonrotating earth, as well as in the realistic case when dealing with the rotating earth.

The verifications connected with the orbital part of SAGG have been somewhat more complicated. This is caused by the fact that the state vector parameters are always weighted. A perfect recovery of shifts introduced in any weighted parameters is in general impossible. A different approach has therefore been undertaken. First, a nonrotating earth has been considered, and then the effect of the earth's rotation has been introduced through the angular velocity ω. Verifications of the orbital part of SAGG have been carried out with a new program, called "Mini - SAGG", created exclusively for this purpose (the final version of this program allows for an arbitrary value of ω). The mathematical model in Mini-SAGG uses formulas completely independent of those in SAGG. This method is called the "analytic F and G series evaluation", and it has been programmed and documented at DBA Systems, Inc.

However, this new approach can be used only in conjunction with Keplerian orbits. The following procedure has therefore been adopted. The earth has been represented by a homogeneous rotating sphere, giving rise to Keplerian orbits. Simulated altimetry data on such orbits have been adjusted independently by Mini-SAGG. The equivalence of both approaches has been achieved by setting, in SAGG, all the C's and S's to zero (and r_0 to a prescribed value) and holding them fixed. The orbits have been varied in inclination and eccentricity. In one test, a circular orbit on the equator has been generated and analysed. The a-priori sigmas of the state vector components have also been varied. In addition, certain known errors have been introduced into the theoretical values of satellite altimetry events to assess their recovery by the state vector parameters.

Besides accomplishing their own goal, these tests have also led to several improvements in the program SAGG. One such improvement results in a better economy. In particular, the run-time has been shortened by introducing analytical formulas rather than inverting certain (6 x 6) matrices numerically, as will be explained next.

It is well known that the (6 x 1) state vectors can be transformed between the inertial and the E.F. systems by means of certain (6 x 6) matrices as follows:

$$(\text{state vector})_{\text{E.F.}} = \bar{R} \times (\text{state vector})_{\text{inertial}} ,$$

and

$$(\text{state vector})_{\text{inertial}} = \tilde{R} \times (\text{state vector})_{\text{E.F.}} ,$$

where, necessarily,

$$\tilde{R} = \bar{R}^{-1} \quad . \tag{4.24}$$

If the chosen inertial system is made to coincide with the E.F. system at t_0, we have

$$\bar{R} = \begin{bmatrix} I & 0 \\ T^T & I \end{bmatrix} \quad , \tag{4.25}$$

where T has been given in (4.9b). If (4.25) is inverted analytically, one obtaines

$$\tilde{R} = \begin{bmatrix} I & 0 \\ T & I \end{bmatrix} \quad , \tag{4.26}$$

which is in fact the (6 x 6) matrix shown in (4.9b).

It follows from the above argument that if $\bar{R}$ is available, its inversion in (4.24) can be written down immediately as (4.26). In SAGG, $\bar{R}$ is computed first; it is needed e.g. for the formation of the variance-covariance matrix of the state vector components in the E.F. system rather than in the chosen inertial system in which these components are originally given. The matrix $\tilde{R}$ is needed later in computing $\tilde{M}_1$. It is noteworthy that Mini-SAGG can be -- and indeed has been -- utilized when an adjustment is taking place in coordinate systems other than the E.F. system. In particular, an option to adjust in the E.F. system or in a (general) chosen inertial system may be exercised.

This concludes the final verification of SAGG with respect to its orbital part which had not been tested previously. All the final results and variances-covariances agree with an independent approach to all the significant digits printed, thus verifying the theory and the program.

4.4 Adjustment of Short Arc Orbits when the Geoid is Considered Fixed

Suppose that the geoid surface has been well established over a given area upon adjusting hundreds or perhaps even thousands of arcs of satellite altimetry. It may then become desirable, in practice, to analyse in some efficient way one or more additional arcs in that area. If we can reasonably assume that the geoidal parameters will not change due to this new -- but weak -- information, the computation of residuals for such isolated arcs will be an easy matter. The residuals may subsequently serve in an orbital analysis and/or as a means to express certain detailed geoidal features along the given profiles.

The observation equations in this mode closely resemble those of Section 4.1, except that the corrections to the potential coefficients are now identically zero. In fact, due to the geoid being considered fixed, an almost identical approach could be used also in conjunction with representations of the geoid surface other than the spherical harmonics. Upon deleting the geoidal parameters from (4.7), the observation equations are symbolized by

$$
\begin{aligned}
V = \{ & (X_s/R) \, | \, 1 - (a/R)\, e^2 \sin^2 \bar{\phi} \, | \quad , \\
& (Y_s/R) \, | \, 1 - (a/R)\, e^2 \sin^2 \bar{\phi} \, | \quad , \\
& (Z_s/R) \, | \, 1 + (a/R)\, e^2 \cos^2 \bar{\phi} \, | \, \} \\
& \times \tilde{M}_1 \begin{bmatrix} dX \\ dY \\ dZ \\ \dot{dX} \\ \dot{dY} \\ \dot{dZ} \end{bmatrix} + L \quad , \qquad (4.27)
\end{aligned}
$$

where L is given as in (4.8), except that the symbol r^0 should be understood as "r^{fixe}", computed in this case by (4.1) or (4.4) with the fixed geoidal parameters. In analogy to (4.8) and the explanation that followed, we may write

$$L = R^0 - H^b + d - r^{fixe} \quad , \tag{4.28a}$$

or

$$L = (X_p^2 + Y_p^2 + Z_p^2)^{\frac{1}{2}} - r^{fixe} \quad . \tag{4.28b}$$

In this type of adjustment, the state vector parameters are made to fit, in the best possible way, the geoidal surface along a particular short arc.

Equation (4.27) may be understood as representing either one observation equation or a set of observation equations. These two cases may be written respectively as

$$\underset{(1,1)}{v_i} = \underset{(1,6)}{a_i}\ \underset{(6,1)}{X} + \underset{(1,1)}{l_i} \quad , \tag{4.29a}$$

$$\underset{(n,1)}{V} = \underset{(n,6)}{A}\ \underset{(6,1)}{X} + \underset{(n,1)}{L} \quad , \tag{4.29b}$$

where the symbols are easily identified (the number of satellite altimetry observations H^b along a given arc is denoted as n). The normal equations are formed from (4.29b) in a standard way:

$$(A^T P A + P_X) X + A^T P L = 0 \quad , \tag{4.30}$$

where P is the weight matrix associated with the observations (H^b).

The observations are often independent and of an equal precision (the precision is represented by σ, the standard error). We then have

$$A^T P A = A^T A / \sigma^2 \quad , \tag{4.31a}$$

$$A^T P L = A^T L / \sigma^2 \quad . \tag{4.31b}$$

The symbol P_X denotes the weight matrix associated with the state vector parameters as obtained from (3.26). Such a weighting process makes the matrix of normal equations nonsingular; accordingly, from (4.30) and (4.31) we deduce that

$$X = - (A^T A / \sigma^2 + P_X)^{-1} A^T L / \sigma^2 \quad , \tag{4.32}$$

or

$$\underset{(6,1)}{X} = \underset{(6,6)}{- N^{-1}} \quad \underset{(6,1)}{U} \quad .$$

Since every arc is adjusted independently, only one inversion of a (6 x 6) matrix is needed in each adjustment. The matrix A^TA and the vector A^TL can be formed efficiently as

$$\underset{(6,6)}{A^T A} = \sum_{i=1}^{n} \underset{(6,1)}{a_i^T} \quad \underset{(1,6)}{a_i} \quad ,$$

$$\underset{(6,1)}{A^T L} = \sum_{i=1}^{n} \underset{(6,1)}{a_i^T} \quad \underset{(1,1)}{l_i} \quad .$$

After solving for the parameters X, each residual can be computed seperately from (4.29a). It follows that only the matrices and vectors of very small dimensions need to be stored simultaneously. In fact, the largest matrix present in the computer core has the dimensions (6 x 6). The computer run-time requirements for processing even a large number of arcs in this manner are also relatively very small; this is attributable to the fact that the only inversions needed are those of one (6 x 6) matrix per arc. We may thus conclude that the above computations could be performed easily and efficiently even on small computers.

5. GRAVITY ANOMALY ADJUSTMENT MODEL

5.1 Usual Formula for Gravity Anomalies in Terms of the Spherical Harmonic Potential Coefficients

In this section, we present a well-known formula relating approximate gravity anomalies ($\Delta\tilde{g}$) to the spherical harmonic potential coefficients. In the subsequent sections, this formula will be re-derived, interpreted, and analysed. For it is important to know to what extent approximations are involved in the gravity anomaly model, later to be used in a combined adjustment with the satellite altimetry model.

The formula that has been selected as a simple, yet reasonably accurate basis for adjustments of the gravity-type data is the following:

$$\Delta\tilde{g} = (kM/r^2) \sum_{n=2}^{\infty} (n-1)(a/r)^n \sum_{m=0}^{n} (\Delta C_{nm} \cos m\lambda + \Delta S_{nm} \sin m\lambda) P_{nm}(\sin\bar{\phi}), \quad (5.1)$$

where r, $\bar{\phi}$, and λ are the spherical coordinates of a geoidal point P, kM is the product of the gravitational constant and the earth's mass, $P_{nm}(\sin\bar{\phi})$ are the associated Legendre functions (here not normalized), and "a" is the equatorial radius of the earth. Further, we have

$$\Delta C_{20} = C_{20} - C^*_{20} \quad ,$$
$$\Delta C_{40} = C_{40} - C^*_{40} \quad ,$$
$$\Delta C_{60} = C_{60} - C^*_{60} \quad ,$$

the other ΔC's and ΔS's being essentially the conventional spherical harmonic potential coefficients (C's and S's) themselves. The values C^*_{20}, C^*_{40}, and C^*_{60}

refer to the equipotential ellipsoid of revolution adopted as a base reference surface in the normal gravity field. In the actual field, a corresponding reference surface is the geoid.

The impact of replacing the coefficients C^*_{80}, etc., by zeros is negligible. Already, the maximum possible contribution of $C^*_{60} \simeq -0.000\,000\,0061$ in equation (5.1) represents merely 0.03 mgal (it also has a small effect on the potential of the ellipsoid). The contribution of C^*_{80} would be two orders of magnitude smaller and a similar pattern would hold for further coefficients. By comparison, the maximum contribution of $C^*_{40} \simeq +0.000\,0024$ is 7 mgal. As a matter of interest, we mention that similar characteristics would be found with regard to the spherical harmonic expansion of γ_r (radial component of normal gravity); the numerical values would be of course slightly different. In particular, associated with C^*_{60} and C^*_{40}, one would find 0.04 mgal and less than 12 mgal, respectively.

A relation similar to (5.1) had been previously compared by Rapp [1972'], equation (3), with less accurate formulas as well as with the rigorous results $g - \gamma$. We propose to carry the latter comparisons one step further. In particular, we shall demonstrate that (5.1) embodies two approximations, both related to the familiar spherical approximation. To distinguish them clearly in the text, one will be called "direction approximation" and the other, "spherical approximation". They will be analysed numerically in conjunction with different sets of potential coefficients and for the geodetic coordinates ranging over the whole globe.

5.2 Gravity Components, Gravity Disturbance, and Gravity Anomaly in the Radial Direction

We begin by recalling a few definitions associated with the normal direction to the geoid or to the reference ellipsoid, and by creating some new definitions associated with the radial direction; the latter case will be identified by the subscript "r". When dealing with the normal field, it will be necessary to distinguish between the quantities (unprimed) pertaining to the ellipsoidal point Q and the quantities (primed) pertaining to the ellipsoidal point Q'. The normal gravity (γ) referring to the geoidal point P will have to be distinguished by the subscript "P" from the normal gravity referring to either of the ellipsoidal points; thus, in one particular case (when γ indicates the normal gravity at point P in the radial direction) we shall employ the double subscript "r, P". The quantities in the actual field referring to the geoidal point P will not bear the subscript "P". Figure 5.1 shows the configuration of points and some related quantities in a meridian plane. This figure bears a close resemblance to Figure 4.1.

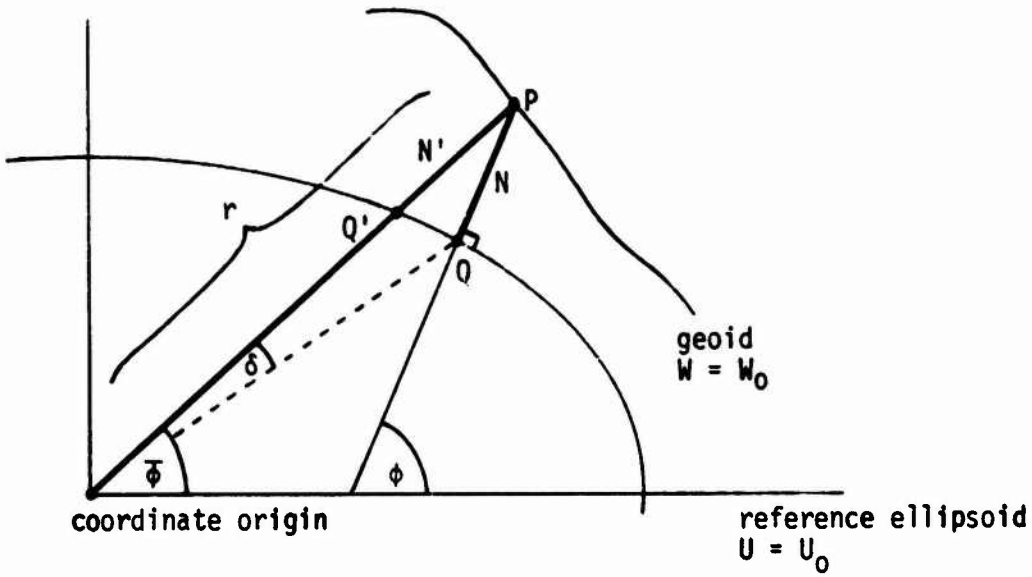


FIGURE 5.1

Geoidal point P and the corresponding points Q, Q' on the reference ellipsoid.

The actual, normal, and disturbing potential at an arbitrary point P in space are denoted as W, U_P, and T, respectively. According to customary conventions, we define

$$W = U_P + T \tag{5.2}$$

and consider it at the geoidal point P of Figure 5.1. In this equation, U_P may be expressed either as a function of N (the usual way), or as a function of N' (the specific way of this analysis). Since according to Chapter 1 we consider the reference ellipsoid to have the same potential ($U = U_0$) as the geoid, we have

$$U_0 = W_0 \quad . \tag{5.3}$$

Bearing in mind that the radial distance increases "upwards" while the gravity increases "downwards", we notice the sign convention in the formulas

$$g_r = - \partial W/\partial r \quad \text{... at point P} \quad , \tag{5.4a}$$

$$\gamma_r = - \partial U/\partial r \quad \text{... at point Q} \quad , \tag{5.4b}$$

$$\gamma'_r = - (\partial U/\partial r)' \quad \text{... at point Q'} \quad . \tag{5.4c}$$

From (5.2), we express in conjunction with (5.4c) and (5.3):

$$W_0 = U_0 - \gamma'_r N' + T \quad ,$$

$$N' = T / \gamma'_r \quad . \tag{5.5}$$

Equation (5.5) could be called "specific Bruns formula" whose more usual form appears in [Heiskanen and Moritz, 1967], page 85; henceforth this reference will be abbreviated as [HM].

The usual definitions of the gravity disturbance and the gravity anomaly are respectively

$$\delta g = g - \gamma_p \quad , \tag{5.6a}$$

$$\Delta g = g - \gamma \quad . \tag{5.6b}$$

We now define

$$\delta g_r = g_r - \gamma_{r,P} \quad , \tag{5.7a}$$

$$\Delta g_r = g_r - \gamma'_r \quad . \tag{5.7b}$$

Following an approach similar to that of [HM], pages 85, 86, we find

$$\begin{aligned} \gamma_{r,P} &= \gamma'_r + (\partial\gamma_r / \partial r)' \; N' \quad , \\ g_r - \gamma'_r &= g_r - \gamma_{r,P} + (\partial\gamma_r / \partial r)' \; N' \quad , \\ \Delta g_r &= \delta g_r + (\partial\gamma_r / \partial r)' \; N' \quad , \end{aligned} \tag{5.8}$$

upon taking equations (5.7) into account. Recalling (5.4a) and using an analogous equation in the normal gravity field, namely

$$\gamma_{r,P} = - (\partial U / \partial r)_P \quad ,$$

from (5.7a) we can write

$$\delta g_r = - \partial(W - U_P) / \partial r \quad ,$$

or

$$\delta g_r = - \partial T / \partial r \quad . \tag{5.9}$$

Considering (5.5) and (5.9), from (5.8) we finally obtain

$$\Delta g_r = - \partial T / \partial r + [(1 / \gamma_r)(\partial\gamma_r / \partial r)]' \; T \quad , \tag{5.10}$$

which is an exact form of the boundary condition on the geoid. This form is particularly suited for the analysis of errors entailed by equation (5.1).

The quantity Δg_r in the boundary condition (5.10) was defined in (5.7b). However, this definition may be slightly modified, affecting the value of Δg_r by only a completely negligible amount. In particular, we replace γ' by γ and find what change it entails in terms of milligals. As in (4.12), the angle δ separating the ellipsoidal points Q and Q' is computed as

$$\delta \simeq \tfrac{1}{2} e^2 (N / R) \sin 2\phi \quad ,$$

and thus

$$| d\phi |_{max} \simeq | \delta |_{max} \simeq \tfrac{1}{2} e^2 (N / R) \quad .$$

Even if the geoid undulation were as high as 120 m, we would have

$$| d\phi |_{max} \simeq 0.013" \quad .$$

On the other hand, differentiating the formula giving γ as a function of latitude, from [HM], page 79, we have to a good approximation:

$$d\gamma \simeq 978 \text{ gal} \times 0.0053 \sin 2\phi \, d\phi \quad .$$

Even if we took $d\phi = 0.013"$ and $\sin 2\phi = 1$, we would obtain

$$d\gamma \simeq 0.0003 \text{ mgal} \quad ,$$

which proves that we can indeed write $\gamma' = \gamma$, $\gamma'_r = \gamma_r$, and thus

$$\Delta g_r = g_r - \gamma_r \quad , \qquad (5.11)$$

replacing the definition (5.7b).

To express gravity anomalies in terms of spherical harmonic coefficients, we introduce two approximations. The first, "direction" approximation, amounts to replacing Δg_r by Δg. The second, "spherical" approximation, has been used e.g. in [HM], page 87, in connection with the anomaly field; it emerges when one accepts the relation

$$\gamma \quad = \quad kM / r^2 \quad ,$$

which would be exact if the reference ellipsoid were a nonrotating homogeneous sphere, centered at the coordinate origin. If, in addition, the geoid (no density assumption is involved) were assumed to coincide with this sphere, the points Q, Q' would then coincide with P, separated from the origin by a distance r. With such simplification, we would also have $\gamma_r \equiv \gamma$ and

$$\left[(1 / \gamma_r)(\partial\gamma_r / \partial r)\right]' \equiv (1 / \gamma)(\partial\gamma / \partial r) = - 2 / r \quad .$$

Upon using these two approximations in (5.10), we obtain

$$\Delta\tilde{g} \quad = \quad - \partial T / \partial r \; - \; (2 / r)\, T \quad , \qquad (5.12)$$

where we have anticipated the result to correspond to equation (5.1).

In terms of $\Delta\tilde{g}$, the rigorous gravity anomaly (Δg) may be expressed as follows:

$$\Delta g \quad = \quad \Delta\tilde{g} \; + \; \Delta_{tot} \quad , \qquad (5.13a)$$

$$\Delta_{tot} \quad = \quad \Delta_1 \; + \; \Delta_2 \quad , \qquad (5.13b)$$

$$\Delta_1 \quad = \quad \Delta g \; - \; \Delta g_r \quad , \qquad (5.13c)$$

$$\Delta_2 \quad = \quad \Delta g_r \; - \; \Delta\tilde{g} \quad . \qquad (5.13d)$$

The quantity Δ_{tot}, which in fact is

$$\Delta_{tot} = \Delta g - \Delta\tilde{g} \quad , \tag{5.13b'}$$

is called the total error; here it has the nature of a correction (added to $\Delta\tilde{g}$, it yields Δg). Similarly, Δ_1 is called the direction error and Δ_2 is called the spherical error, and they result from the two approximations described in the preceding paragraph. From (5.6b) and (5.11), we also have

$$\Delta_1 = (g - g_r) - (\gamma - \gamma_r) \quad . \tag{5.13c'}$$

We shall next express some of the above quantities explicitly.

5.3 Series in Terms of the Spherical Harmonic Potential Coefficients

We begin by recalling the formula for the potential (W) in the actual gravity field of the earth. It will be used to verify that equation (5.12) indeed corresponds to (5.1), confirming thus the breaking down of Δg as indicated in equations (5.13). The same formula will subsequently serve to express the three components of the gravity vector which, in turn, will enable us to compute all the desired quantities in (5.13). In terms of the conventional C's and S's, the potential in space (but at points rotating with the earth) is given as

$$W = (kM/r)\left[1 + \sum_{n=2}^{\infty}(a/r)^n \sum_{m=0}^{n}(C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)P_{nm}(\sin\bar{\phi})\right] + \tfrac{1}{2}\omega^2 r^2\cos^2\bar{\phi}\ , \quad (5.14)$$

provided of course that it is evaluated outside the masses. In the normal gravity field, the potential in space is similarly given by the formula

$$\begin{aligned} U = (kM/r)[1 &+ C^*_{20}(a/r)^2 P_2(\sin\bar{\phi}) + C^*_{40}(a/r)^4 P_4(\sin\bar{\phi}) \\ &+ C^*_{60}(a/r)^6 P_6(\sin\bar{\phi})] + \tfrac{1}{2}\omega^2 r^2\cos^2\bar{\phi}\ , \end{aligned} \quad (5.15)$$

where the terms with C^*_{80}, etc., have been neglected in agreement with our previous discussion. Comparing this expression with [HM], page 73, we identify:

$$\begin{aligned} C^*_{20} &= -J_2\ , \\ C^*_{40} &= -J_4\ , \\ C^*_{60} &= -J_6\ . \end{aligned}$$

From (5.14) and (5.15), it follows that in agreement with Chapter 1, the same mass (M) has been assumed in the actual and in the normal gravity fields. A similar assumption holds true also with respect to the rotational velocity (ω). The term $\frac{1}{2}\omega^2 r^2 \cos^2\bar{\phi}$ implies that

$$C_{21} = S_{21} = 0 ,$$

the Cartesian z-axis thus being parallel to the earth's axis of rotation. Since C_{10}, C_{11}, S_{11} are missing in (5.14), the earth's center of mass and the Cartesian origin coincide.

In considering (5.2) together with (5.14) and (5.15) at an external point, the distrubing potential in space is seen to be

$$T = (kM/r)\sum_{n=2}^{\infty}(a/r)^n \sum_{m=0}^{n}(\Delta C_{nm}\cos m\lambda + \Delta S_{nm}\sin m\lambda)P_{nm}(\sin\bar{\phi}), \quad (5.16)$$

where ΔC's and ΔS's were identified following (5.1). Upon using it with (5.12), derived for the geoid surface, we recover equation (5.1). For the procedure to be valid, the masses outside the geoid must be considered absent. Mathematically, this is achieved by a suitable gravity reduction, see e.g. [HM], Chapter 3. We have seen that the (precise) boundary condition (5.10) implied the same potential of the reference ellipsoid and the geoid, and thus the same value of r_0 (a quantity to be defined subsequently). If we adopt the same J_2 for the normal field and the actual field, the reference ellipsoid is then the mean earth ellipsoid, since M and ω have been already assigned the same numerical values in either field. The formula (5.1) is unchanged whether the ellipsoid is in absolute position (C_{10}, C_{11}, S_{11} missing) or not, due to the factor (n - 1).

Our next task is to express the magnitude (g) of the gravity vector $\vec{g}$ = grad W on the geoid, whose three orthogonal components are g_r, $g_{\bar{\phi}}$, g_λ. These components are associated with the directions of the three coordinate lines in the (geocentric) spherical coordinate system; the differential distances along these coordinate lines are respectively $d\ell_r = dr$, $d\ell_{\bar{\phi}} = r d\bar{\phi}$, and $d\ell_\lambda = r \cos\bar{\phi}\, d\lambda$. Keeping in mind the sign convention seen in (5.4a), we express

$$g_r = - \partial W / \partial r \quad , \tag{5.17a}$$

$$g_{\bar{\phi}} = (1 / r)\ \partial W / \partial \bar{\phi} \quad , \tag{5.17b}$$

$$g_\lambda = \left| 1 / (r \cos \bar{\phi}) \right| \partial W / \partial \lambda \ . \tag{5.17c}$$

For the given $\bar{\phi}$, λ coordinates, an iterative solution for the radial distance (r) from the coordinate origin to the geoid was presented in [Blaha, 1975], equation (46a), as

$$r = r_0 \left[1 + \sum_{n=2}^{\infty} (a/r)^n \sum_{m=0}^{n} (C_{nm} \cos m\lambda + S_{nm} \sin m\lambda) P_{nm}(\sin\bar{\phi}) \right] + \tfrac{1}{2}\, \omega^2 r_0 r^3 \cos^2\bar{\phi} / (kM) \quad . \tag{5.18}$$

In this expression, r_0 corresponds to the radius of a ficticious spherical earth having the same potential as the geoid, namely

$$r_0 = kM / W_0 \ ; \tag{5.19a}$$

due to (5.3), we can write

$$r_0 = r_0^* \quad , \tag{5.19b}$$

where in the normal field, r_0^* is the counterpart of r_0. The equality in

(5.19b) was already indicated in the preceding paragraph. Upon differentiating (5.14) according to equations (5.17), one obtains the standard formulas below:

$$g_r = (kM/r^2)\left[1 + \sum_{n=2}^{\infty}(n+1)(a/r)^n \sum_{m=0}^{n} (C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)P_{nm}(\sin\bar{\phi})\right] - \omega^2 r\cos^2\bar{\phi}\ , \tag{5.20a}$$

$$g_{\bar{\phi}} = (kM/r^2)\sum_{n=2}^{\infty}(a/r)^n \sum_{m=0}^{n} (C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)dP_{nm}(\sin\bar{\phi})/d\bar{\phi} - \omega^2 r\sin\bar{\phi}\cos\bar{\phi}\ , \tag{5.20b}$$

$$g_\lambda = [kM/(r^2\cos\bar{\phi})]\sum_{n=2}^{\infty}(a/r)^n \sum_{m=0}^{n} m(-C_{nm}\sin m\lambda + S_{nm}\cos m\lambda)P_{nm}(\sin\bar{\phi}). \tag{5.20c}$$

The required gravity value is then computed as

$$g = (g_r^2 + g_{\bar{\phi}}^2 + g_\lambda^2)^{1/2}\ . \tag{5.21}$$

5.4 Numerical Results

The notion of the mean earth ellipsoid (MEE) was introduced e.g. in [HM], page 109. In the present analysis, we shall be using the Geodetic Reference System 1967 (GRS 1967). In this system, the MEE shares with the actual earth the four "defining parameters", kM, ω, a, and C_{20} (alternatively, U_0 could replace a, but only four parameters are independent). The other "derived parameters" of the MEE, such as C^*_{40}, C^*_{60}, r^*_o, e^2, γ_e, γ_p are uniquely determined from the defining parameters. We present below the numerical values of all the parameters used, including the normalized C-coefficients in parentheses.

1) Defining parameters:

$kM = .3986030 \times 10^{15}\ m^3/sec^2$,

$\omega = .72921151 \times 10^{-4}$ rad/sec ,

$a = 6378160.0$ m ,

$C^*_{20} = -1.08270 \times 10^{-3} = C_{20}$ $\quad (\bar{C}^*_{20} = -.48419816 \times 10^{-3} = \bar{C}_{20})$.

2) Derived parameters:

$C^*_{40} = 2.37126 \times 10^{-6}$ $\quad (\bar{C}^*_{40} = .790420 \times 10^{-6})$,

$C^*_{60} = -6.0852 \times 10^{-9}$ $\quad (\bar{C}^*_{60} = -1.68773 \times 10^{-9})$,

$r^*_o = 6363695.672\ m = r_o$,

$e^2 = .00669460533$,

$\gamma_e = 978031.8456$ mgal ,

$\gamma_p = 983217.7279$ mgal .

The reference values C^*_{40} and C^*_{60} were derived from the formula (2-92) of [HM], which may be rewritten in terms of C^*_{20} as

$$J_{2n} = (-1)^{n+1}\, 3e^{2n}\, (1-n+5nJ_2/e^2) \,/\, \left[(2n+1)(2n+3)\right] \quad ,$$

where $J_{2n} \equiv -C^*_{2n,\,0}$; the normalized value of a coefficient (k, 0) is of course obtained upon dividing the conventional value by $(2k+1)^{\frac{1}{2}}$. As another example, in analogy to equations (5.19) we have

$$r^*_o = kM \,/\, U_o \quad ,$$

where

$$U_o = 6.2637030523 \times 10^7 \ m^2 \,/\, sec^2 \quad ,$$

U_o being in this case a derived parameter. Due to their uncertainties, several of the above values are used in practice to fewer significant digits than indicated. However, in order to maintain the highest possible precision in the analysis, we wish to work with the parameters that are almost perfectly consistent, in the sense that the round-off errors in their values lead to discrepancies significantly smaller than 0.001 mgal.

That the presented values indeed satisfy this requirement may be illustrated as follows. A program has been created in which g on the geoid is computed from a set of spherical harmonic potential coefficients (C's and S's). Besides these coefficients, the other parameters involved are kM, ω, and r_o. If one works in the GRS 1967, these three parameters as well as C_{20} take on the values listed above. On the other hand, the normal gravity (γ) is computed by the rigorous formula of Somigliana, which involves the

parameters a, e^2, γ_e, and γ_p of the MEE. It reads

$$\gamma = (a\gamma_e \cos^2\phi + b\,\gamma_p \sin^2\phi)/(a^2 \cos^2\phi + b^2 \sin^2\phi)^{\frac{1}{2}} \quad , \quad (5.22)$$

where $b = a(1 - e^2)^{\frac{1}{2}}$ is the semi-minor axis of the reference ellipsoid, γ_e and γ_p are the values of normal gravity at the equator and at the poles, respectively, and ϕ is the geodetic latitude (see e.g. [HM], page 70). If we set all the C's and S's to zero with the exception of C_{20} (it is equal to C^*_{20}), C_{40}, and C_{60}, and if we set further $C_{40} = C^*_{40}$, $C_{60} = C^*_{60}$, we have made the "geoid" coincide almost perfectly with the MEE; the only conceivable discrepancy may be due to the neglected coefficients C^*_{80}, etc., which are extremely small. Under such circumstances the value of g computed from equations (5.18), (5.20), (5.21) should be equal to γ. The outcome of this test has been encouraging. All the deviations from the anticipated results have been listed as 0.000 mgal, not only for g itself, but also for two of its three orthogonal components expressed numerically, with ϕ, λ ranging over the whole globe. Confidence has thus been gained in the GRS 1967 values to the number of significant digits printed, in that a detailed and accurate numerical analysis may be performed, reliable to better than 0.001 mgal.

We next present the formulas used in computing Legendre polynomials, associated Legendre polynomials, and their first derivative with respect to the geocentric latitude ($\bar{\phi}$). The numerical values of these functions are used in equations (5.20) in conjunction with (5.18). We begin by listing the few notation conventions, in which $x \equiv \sin \bar{\phi}$:

$P_n \equiv P_n(x) \equiv P_{n0}(x)$ ... Legendre polynomial in argument x,

$P_{nm} \equiv P_{nm}(x)$ ... Associated Legendre polynomial in argument x,

$p_n^{[m]} \equiv d^m P_n / dx^m$.

The recurrence formulas that have been found useful and efficient computationally are

$$P_n = (1/n)\left[x\,(2n-1)P_{n-1} - (n-1)P_{n-2}\right], \qquad n \geqslant 1;$$

$$p_n^{[1]} = p_{n-2}^{[1]} + (2n-1)P_{n-1} , \qquad (m=1),$$

$$p_n^{[m]} = p_{n-2}^{[m]} + (2n-1)p_{n-1}^{[m-1]} , \qquad 2 \leqslant m \leqslant n-2 ,$$

$$p_n^{[m]} = (2n-1)p_{n-1}^{[m-1]}, \qquad m = n-1, n ;$$

the last three equations lead to

$$P_{nm} = (1-x^2)^{m/2}\, p_n^{[m]} \equiv \cos^m \bar{\phi}\; p_n^{[m]} .$$

The recurrence formulas that appear simplest and most efficient for the subsequent computation of the required first derivatives are

$$dP_n / d\bar{\phi} = P_{n1} , \quad n \geqslant 1 ;$$

$$dP_{nm} / d\bar{\phi} = -m\, tg\bar{\phi}\; P_{nm} + P_{n,\,m+1} , \quad 1 \leqslant m \leqslant n-1 ,$$

$$dP_{nn} / d\bar{\phi} = -n\, tg\bar{\phi}\; P_{nn} , \quad (m = n) .$$

Due to $tg\,\bar{\phi}$ in these formulas, $\bar{\phi}$ is kept from approaching 90^0 closer than 0.0001" in order to avoid numerical problems. As an alternative expression

to the last two equations, one could use

$$dP_{nm} / d\bar{\phi} = \cos^{m-1} \bar{\phi} \left[m \sin \bar{\phi} \, P_n^{[m]} - (n-m+1)(n+m) P_n^{[m-1]} \right] .$$

The numerical agreement between the corresponding results has been perfect.

We may now present the results of various numerical tests. The final quantities listed in Table 5.1 are Δ_1, Δ_2, and Δ_{tot}; they appear under the headings "dir.", "sph.", and "tot.", and are printed to the nearest 0.001 mgal. All the computational formulas have been already given (g is computed from 5.21, etc., γ from 5.22, Δg from 5.6b, $\Delta \tilde{g}$ from 5.1, Δg_r from 5.11, Δ_1 from 5.13c or 5.13c', Δ_2 from 5.13d, and Δ_{tot} from 5.13b or 5.13b'). We mention that $\gamma - \gamma_r$ is computed with advantage from

$$\gamma - \gamma_r = f(a, J_2, \ldots, \bar{\phi}) \; ; \qquad (5.23)$$

its explicit formulation, reliable to better than 0.001 mgal for an arbitrary level ellipsoid, is given in Appendix 2, equations (A2.10), (A2.11). Consequently, γ_r is computed from

$$\gamma_r = \gamma - f(a, J_2, \ldots, \bar{\phi}) \quad .$$

The geoid is generated by a set of reasonable potential coefficients, complete through the degree and order (20, 20). This set is represented here by the "synthetic potential coefficients" supplied by AFGL (the coefficients are normalized). They have been unaltered except for the coefficients (2, 1) which have been set precisely to zero in agreement with our previous statement, and the coefficient (2, 0) which has been set precisely equal to the GRS 1967 value; in any event, these changes are extremely small. The coefficient (1, 0)

as well as the coefficients (1, 1) have been absent from the beginning. For the sake of comparison and illustration of the influence of higher order and degree coefficients, two additional cases are listed; they are represented by the same set of potential coefficients as above, but truncated this time to a (12, 12) model and to a (6, 6) model, respectively. The results appear side by side with those of the (20, 20) model. In Table 5.1, the results are presented in such a way that at least one of the three errors (direction, spherical, total) in the (20, 20) model reaches, in absolute value, 0.130 mgal or more. This table could be of course extended if we wanted to list some additional errors smaller than 0.130 mgal, or if we wanted to include some additional models (e.g. 15, 15, or 16, 16, etc.), but little further insight would be acquired. For instance, the table would have 63 rows instead of 47 if the smallest errors listed included 0.125 mgal, but most of its vital characteristics would be preserved; if these errors included 0.120 mgal, the number of rows would be 83.

A discussion of the results from Table 5.1 is in order. We realize that the listed errors depend on the anomalous field, not on the MEE; for if the "geoid" were made to coincide with the MEE, we would have $g = \gamma$, $g_r = \gamma_r$, $\Delta g = 0$, $\Delta \tilde{g} = 0$ (all ΔC's and ΔS's would be zero), and thus identically $\Delta_1 = \Delta_2 = \Delta_{tot} = 0$. The largest total error in the (20, 20) model reaches 0.194 mgal, due entirely to the spherical error; it occurs at the location $\phi = 0^0$, $\lambda = 75^0$ (in a $5^0 \times 5^0$ geographic grid). This value agrees quite well with the maximum (total) error of 0.171 mgal reported by Rapp [1972']. We observe that the spherical errors are usually larger than the direction errors, especially in equatorial regions. A few larger direction errors

		Model (20, 20)			Model (12, 12)			Model (6, 6)		
ϕ^0	λ^0	dir.	sph.	tot.	dir.	sph.	tot.	dir.	sph.	tot.
-40	165	.130	-.019	.111	.087	-.017	.069	.073	-.017	.057
	170	.140	-.013	.127	.096	-.014	.082	.075	-.015	.060
-30	30	-.065	-.067	-.131	-.052	-.060	-.112	-.048	-.063	-.111
	35	-.060	-.076	-.136	-.054	-.069	-.123	-.055	-.065	-.120
-25	30	-.085	-.052	-.137	-.062	-.051	-.113	-.046	-.058	-.104
	35	-.089	-.063	-.152	-.067	-.060	-.127	-.054	-.058	-.112
	40	-.076	-.065	-.141	-.069	-.062	-.131	-.061	-.053	-.113
-10	65	-.032	.148	.116	-.034	.139	.105	-.024	.131	.107
	70	-.034	.152	.118	-.033	.148	.115	-.024	.144	.120
	75	-.035	.146	.112	-.030	.145	.114	-.022	.147	.125
	80	-.037	.134	.098	-.027	.133	.106	-.019	.139	.119
- 5	60	-.007	.151	.143	-.010	.147	.137	-.008	.136	.128
	65	-.006	.176	.170	-.011	.173	.163	-.008	.157	.149
	70	-.007	.183	.176	-.011	.183	.171	-.008	.170	.161
	75	-.011	.182	.172	-.012	.178	.167	-.008	.171	.163
	80	-.014	.175	.161	-.012	.165	.153	-.008	.162	.154
	85	-.016	.158	.143	-.012	.147	.135	-.007	.141	.135
	90	-.014	.131	.117	-.011	.123	.112	-.006	.112	.107
	150	.001	-.132	-.131	.003	-.116	-.113	.002	-.115	-.113
0	60	.000	.161	.162	.000	.162	.162	.000	.150	.150
	65	.000	.181	.181	.000	.188	.188	.000	.171	.171
	70	.000	.189	.189	.000	.199	.199	.000	.184	.184
	75	.000	.194	.194	.000	.197	.197	.000	.186	.186
	80	.000	.193	.193	.000	.186	.186	.000	.176	.176
	85	.000	.182	.182	.000	.170	.170	.000	.155	.156
	90	.001	.155	.155	.000	.145	.146	.000	.125	.126
	150	.000	-.130	-.130	.000	-.122	-.122	.000	-.120	-.120
5	60	-.007	.153	.147	-.007	.154	.147	-.001	.150	.148
	65	-.006	.169	.163	-.008	.178	.169	-.002	.171	.169
	70	-.005	.179	.174	-.007	.191	.184	-.001	.184	.183
	75	-.005	.184	.179	-.004	.194	.190	.000	.188	.187
	80	-.005	.186	.181	.000	.190	.190	.001	.180	.181
	85	-.002	.181	.179	.002	.179	.181	.002	.161	.163
	90	.002	.162	.164	.004	.156	.161	.004	.132	.136
	130	-.000	-.136	-.137	-.001	-.120	-.121	-.002	-.105	-.106
	135	-.001	-.130	-.131	-.002	-.130	-.132	-.002	-.114	-.116
10	65	-.018	.148	.130	-.027	.149	.122	-.012	.157	.146
	70	-.009	.164	.155	-.024	.165	.140	-.011	.171	.161
	75	-.007	.170	.163	-.018	.174	.156	-.009	.177	.168
	80	-.010	.169	.159	-.012	.177	.166	-.005	.172	.166
	85	-.008	.168	.160	-.006	.172	.166	-.002	.157	.155
	90	-.003	.157	.154	-.002	.154	.152	.002	.132	.134
15	70	-.025	.143	.118	-.036	.132	.096	-.025	.148	.123
	75	-.018	.152	.134	-.030	.145	.115	-.022	.155	.133
	80	-.018	.150	.132	-.025	.153	.129	-.017	.154	.137
	85	-.018	.149	.131	-.020	.153	.133	-.011	.143	.133
	90	-.012	.143	.130	-.015	.139	.124	-.004	.124	.120

TABLE 5.1

The direction, spherical, and total errors at selected locations

occur mainly in one area around $\phi = -45^0$. The errors of the (12, 12) model resemble quite closely those of the (20, 20) model. The largest disparity (over 0.04 mgal) may be found for $\phi = -40^0$, $\lambda = 170^0$, where the direction error also reaches a maximum. A computer run with a (15, 15) model indicated that such disparities diminish rather rapidly when the order and degree of the model are increased. To illustrate this, we list a few selected results obtained with this (15, 15) model (the disparities can be observed to range from large to small):

ϕ^0	λ^0	dir.	sph.	tot.
- 40	170	.111	- .016	.095,
- 40	165	.102	- .019	.083,
- 25	35	- .086	- .061	- .147,
- 30	30	- .061	- .067	- .127,
- 5	65	- .007	.176	.169,
0	60	.000	.161	.161.

It appears that the order and degree (20, 20) are completely satisfactory for evaluation of the three types of error. The most important conclusions we may draw at this point are that the errors have an erratic nature depending on the behavior of the anomalous field, that the direction and the spherical errors do not tend to cancel out, that the spherical errors are in general larger than the direction errors, and that large spherical errors are much more numerous than large direction errors. It is thus apparent that from the theoretical standpoint, eliminating the spherical error would improve the accuracy of the gravity anomaly mathematical model.

5.5 Qualitative Evaluation of the Gravity Anomaly Mathematical Model

The goal of this analysis has been a qualitative evaluation of the well-known formula (5.1),

$$\Delta\tilde{g} = (kM/r^2)\sum_{n=2}^{\infty}(n-1)(a/r)^n\sum_{m=0}^{n}(\Delta C_{nm}\cos m\lambda + \Delta S_{nm}\sin m\lambda)P_{nm}(\sin\bar{\phi}),$$

relating approximate gravity anomalies ($\Delta\tilde{g}$) to the spherical harmonic potential coefficients. A relation of this type had been previously compared by Rapp [1972'] with less accurate formulas as well as with the rigorous results $g - \gamma$. The outcome of the latter task is reasonably well confirmed by the numerical results presented herein under the name "total errors". However, the analysis has been carried one step further by splitting the total error into two parts, called the direction error and the spherical error, and by identifying them mathematically. This has been made possible through a slightly different derivation of the boundary condition from the one appearing in the standard geodetic literature. One of the important outcomes of the numerical tests has been the identification of the spherical error as the main source of the total error.

The errors described are usually much smaller than 0.2 mgal. Admittedly, they are thus mostly of academic interest. However, they could become significant in the future if precise gravity anomalies (reliable e.g. to better than 0.1 mgal) became available in sufficient numbers and in appropriate locations, especially when considering the spherical error and its systematic trends in certain equatorial regions. Be that as it may, the

above formula for $\Delta\tilde{g}$ has been interpreted and analysed, and its reliability has been illustrated numerically with respect to an errorless standard.

In spite of a certain amount of implied criticism, the main practical outcome of this analysis has been an increased confidence in the formula (5.1). For, considering the error characteristics of the data used in current reduction procedures, an error that hardly ever reaches 0.2 mgal may indeed be deemed negligible. To see this, we only have to notice that the gravity-type data that has been made available for the combined adjustments consists of the mean free-air gravity anomalies in $5^0 \times 5^0$ geographic blocks, whose standard error ranges from a minimum of 1 mgal to a maximum of 31 mgal. An average standard error is estimated to be somewhere between 10 mgal and 20 mgal. We may thus conclude that in the present situation, the adoption of the formula (5.1) as a gravity anomaly mathematical model appears to be completely satisfactory.

5.6 Gravity Anomaly Observation Equations

In agreement with the outcome of the previous sections, the gravity anomaly mathematical model is represented by (5.1). This model can be written as

$$\Delta g = C \sum_{n=2}^{\infty} (n - 1)(a / r)^n \, \Delta S(n) \quad , \tag{5.24a}$$

where the symbol "~" has been dropped, and where the notation

$$\Delta S(n) = \sum_{m=0}^{n} (\Delta C_{nm} \cos m\lambda + \Delta S_{nm} \sin m\lambda) \, P_{nm} (\sin \bar{\phi}) \tag{5.24b}$$

has been introduced. It can be readily verified that

$$\begin{aligned} \sum_{n=2}^{\infty} (n-1)(a/r)^n \, \Delta S(n) \; = \; & \sum_{n=2}^{\infty} (n-1)(a/r)^n \; S(n) \\ & - (a/r)^2 \, C^{*}_{20} \, P_2 (\sin \bar{\phi}) \\ & - 3 \, (a/r)^4 \, C^{*}_{40} \, P_4 (\sin \bar{\phi}) \\ & - 5 \, (a/r)^6 \, C^{*}_{60} \, P_6 (\sin \bar{\phi}) \quad , \end{aligned} \tag{5.25}$$

where the terms with C^{*}_{80}, etc., have been neglected as usual, and where S(n) had been defined previously in (4.3a). In practice, the series is truncated at a suitable degree and order $n = N$.

When forming the partial derivatives of the gravity anomaly with respect to the potential coefficients, the value of r may be considered constant (the errors thus introduced are completely negligible as will be

shown shortly). Furthermore, one notices that

$$d\Delta C_{nm} \equiv dC_{nm} \ ,$$
$$d\Delta S_{nm} \equiv dS_{nm} \ .$$

From (5.24), we thus obtain

$$d\Delta g = C \sum_{n=2}^{\infty} (n - 1)(a / r)^n \sum_{m=0}^{n} (dC_{nm} \cos m\lambda + dS_{nm} \sin m\lambda) P_{nm} (\sin \bar{\phi}). \qquad (5.26)$$

If r were considered variable, the additional contribution to (5.26) would be

$$(d\Delta g) = - C (dr / r) P \ ,$$

where

$$P = \sum_{n=2}^{\infty} (n - 1)(n + 2)(a / r)^n \Delta S(n) \ ,$$

and where, from (4.5),

$$dr/r \simeq \sum_{n=2}^{\infty} (a/r)^n \sum_{m=0}^{n} (dC_{nm} \cos m\lambda + dS_{nm} \sin m\lambda) P_{nm}(\sin\bar{\phi}) \ .$$

If we accept $(a / r)^n \simeq 1$ in order to simplify the process of evaluating the contribution of the variable r, we obtain

$$d\Delta g + (d\Delta g) \simeq C \sum_{n=2}^{\infty} \left[(n-1) - P \right] \sum_{m=0}^{n} (dC_{nm} \cos m\lambda + dS_{nm} \sin m\lambda) P_{nm}(\sin\bar{\phi}), \qquad (5.26')$$

where

$$P' = \sum_{n=2}^{\infty} (n - 1)(n + 2)\, \Delta S(n) \quad .$$

Since, in (5.26'),

$$n - 1 \geqslant 1 \quad ,$$

and since numerical evaluations have indicated, by contrast, that for any location on the globe one has

$$| P' | < 0.001 \quad ,$$

we deduce that r may be indeed considered fixed. Consequently, (5.26) may be considered satisfactory in the task of forming the partial derivatives (the partial derivatives accurate to two or three significant digits are sufficient). As a result, the model is linear in the parameters.

In agreement with the above discussion, from (5.26) we obtain

$$\partial \Delta g / \partial C_{n0} = (n - 1)(a / r)^n P_n (\sin \bar{\phi}) \times C, \qquad (5.27a)$$

$$\partial \Delta g / \partial C_{nm} = (n - 1)(a / r)^n P_{nm} (\sin \bar{\phi}) \cos m\lambda \times C, \qquad (5.27b)$$

$$\partial \Delta g / \partial S_{nm} = (n - 1)(a / r)^n P_{nm} (\sin \bar{\phi}) \sin m\lambda \times C, \qquad (5.27c)$$

where $m > 0$ has been assumed. The observation equations may now be built in a standard fashion as follows:

$$V = \left[\ldots \partial \Delta g / \partial C_{n0} \ldots, \ \ldots \partial \Delta g / \partial C_{nm} \ldots, \ \ldots \partial \Delta g / \partial S_{nm} \ldots \right] \begin{bmatrix} \vdots \\ dC_{n0} \\ \vdots \\ dC_{nm} \\ \vdots \\ dS_{nm} \\ \vdots \end{bmatrix} + L. \qquad (5.28)$$

The "discrepancy" terms L are formed in the usual way as

$$L = \Delta g^0 - \Delta g^b \quad , \tag{5.29}$$

where Δg^0 is computed from the model equations (5.24) in terms of the initial values of the parameters, and Δg^b is the input (or "observed") value of the gravity anomaly.

6. LEAST SQUARES ADJUSTMENT AND THE VARIANCE - COVARIANCE PROPAGATION

6.1 Least Squares Adjustment of Satellite Altimetry and Gravity Anomaly Data

The first step in a least squares adjustment process is characterized by forming the observation equations. In the present analysis, two types of observation equations have been encountered. The satellite altimetry observation equations have been presented in Section 4.1, equation (4.7). The gravity anomaly observation equations have been derived in Section 5.6 and presented, in a compact form, as equation (5.28). The two kinds of observation equations are treated in two corresponding parts of the program SAGG, the satellite altimetry part and the gravity anomaly part. Each of the two parts can be adjusted separately, or they can be merged and adjusted simultaneously. In fact, one option in SAGG provides for a seperate adjustment of gravity anomalies, followed by a separate adjustment of satellite altimetry, which is followed by a combined, simultaneous adjustment of satellite altimetry and gravity anomalies.

A separate adjustment of gravity anomalies is a simple, straightforward matter. One only has to form the normal equations as indicated e.g. on the bottom line, page 27 of [Blaha, 1975], and solve for the corrections to the potential coefficients. These parameters may be optionally weighted in a flexible manner to be described shortly.

A separate adjustment of satellite altimetry data may be performed according to Section 2.5 of [Blaha, 1975]. The part "terrestrial source" would of course be left out in this case. The adjustment process is now much more complicated than the one described above, since in addition to the earth-related parameters (the potential coefficients), six new parameters (the state vector components) are introduced for each arc. These new parameters are always weighted; their weight matrix in the E.F. system of adjustment is formed according to (3.26). As in the case of the gravity anomaly adjustment, the potential coefficients can be weighted. Finally, the simultaneous adjustment of both satellite altimetry and gravity anomalies can be performed exactly as described in Section 2.5 of the above reference (the "terrestrial source" is now included).

The a-posteriori variance-covariance matrices for both kinds of parameters (potential coefficients, state vector components) are available from SAGG, whether a separate or a combined adjustment has been performed. They are computed according to the formulas (30) in [Blaha, 1975]. Of these, the variance-covariance matrix for the potential coefficients, denoted as $\Sigma_{C's,\ S's}$, will be needed in the next section, in the task of propagating the variances-covariances into certain predicted quantities. We shall next explain the weighting of potential coefficients in SAGG and describe its usefulness.

The real runs with limited data have indicated that the resolution of low degree and order potential coefficients from satellite altimetry may be weak. Similar results had been gathered earlier from some realistic simulations involving gravity data. In a previous version of SAGG, all the coefficients were either completely free or were all weighted considering their "sigmas" equal to their absolute value itself (an inefficient option

rarely exercised). From the practical standpoint, four tasks have been accomplished and implemented in both the gravity and the altimetry parts of SAGG:

(a) the coefficients have been optionally weighted up to a desired degree and order,

(b) an option has been provided for weighting of only the zonal coefficients up to a desired degree,

(c) the leading coefficient C_{20} has been allowed to be weighted separately,

and

(d) all the weights have been given a great amount of flexibility, in that the magnitudes of the sigmas have been allowed to vary.

Without such changes, adjustments of real data could sometimes produce highly unrealistic corrections to the coefficients, especially if the data are insufficient.

Finally, we give a brief description of the data handling capabilities in SAGG. The option for using real data in the gravity anomaly part of SAGG was built into the program when this part was first conceived. However, one important task performed during this analysis has consisted in modifying the SAGG program so that it can accept real satellite altimetry data. Previously, the data had been simulated by an orbital integrator routine, generating events on a specified orbit which was then broken into a number of short arcs. After the modifications, the orbital information can also be read from a tape containing the epoch time, the state vector components in a chosen inertial coordinate system, and event times for each arc; there is no limitation on the number of arcs. The adjustment takes place

in the earth fixed (E.F.) coordinate system. As we have mentioned, it is advantageous to choose an inertial system (different for each arc) to coincide with the E.F. system. However, an additional, separate program development has been made in which one common inertial system may be used. This may prove useful if, in the future, the state vector components should be given in the inertial coordinate system referred e.g. to the epoch of 1950.0. Thus, two options for handling real satellite altimetry data can be exercised in SAGG. They have been thoroughly tested and yield completely identical results.

6.2 Predictions and the Variance - Covariance Propagation

The quantities that we shall predict using the adjusted potential coefficients (C's and S's) are the geoid undulations (N) and the gravity anomalies (Δg). Most often, these quantities will be predicted in a grid (e.g., a $5^0 \times 5^0$ geographic grid), which may then be used to construct contour maps. In the case of gravity anomalies, the formulas for predicted values will have exactly the same structure as the corresponding formulas for adjusted observations (Δg^a). On the other hand, the adjusted observations in the satellite altimetry model, H^a, would involve not only the adjusted C's and S's as do the quantities Δg^a, but also the adjusted state vector parameters. They would thus have quite different error propagation characteristics from the predicted geoid undulations that we seek.

In (4.17) and (4.18), it was shown that

$$N \simeq N' = r - r' \quad ; \tag{6.1}$$

for a given $\overline{\phi}$, λ, the value of r can be computed from the potential coefficients as in (2.2), and r' is given by the simple formula (4.23) as a function of the same geocentric latitude $\overline{\phi}$. It has been shown that the error in replacing N by N' is completely negligible. In particular, for N as large as 120 m, it holds that

$$|N' - N| < 0.7 \text{ mm} \quad .$$

Since r' is not a function of adjustable parameters, we have

$$\sigma_N^2 = \sigma_{N'}^2 = \sigma_r^2 \quad . \tag{6.2}$$

The gravity anomaly is defined as

$$\Delta g = g - \gamma \quad , \tag{6.3a}$$

where γ, the normal gravity, is independent of the adjustable parameters. In Chapter 5, we have seen that

$$\Delta g \simeq \Delta g_r \simeq \Delta\tilde{g} \quad . \tag{6.3b}$$

The error introduced by interchanging Δg and Δg_r has been called the direction error. On the other hand, the interchange of Δg_r and $\Delta\tilde{g}$ gives rise to the spherical error. Both these errors have been deemed negligible for the current purposes; Δg is thus given by the model equation (5.24a) which, in fact, has been derived for $\Delta\tilde{g}$. Since by definition we also have

$$\Delta g_r = g_r - \gamma_r \quad , \tag{6.3c}$$

if follows that we can write

$$\sigma^2_{\Delta g} = \sigma^2_g = \sigma^2_{\Delta g_r} = \sigma^2_{g_r} \quad . \tag{6.4}$$

It is clear that the values of both N and Δg can be predicted whether only one model has been adjusted separately (satellite altimetry or gravity anomalies), or both models have been adjusted simultaneously. The formulas for the predictions as well as for the variance-covariance propagation remain the same. For a chosen location (i.e., for a given $\bar{\varphi}$, λ), the predictions for r and Δg are computed respectively from (2.2) and (5.24a), where the adjusted values of the potential coefficients must be utilized. The prediction for N is obtained upon subtracting r', given by (4.23), from r as has been shown in (6.1)

The differential increment is r is given as

$$dr = \left| \ldots \partial r/\partial C_{n0} \ldots, \ \ldots \partial r/\partial C_{nm} \ldots, \ \ldots \partial r/\partial S_{nm} \ldots \right| \begin{bmatrix} \vdots \\ dC_{n0} \\ \vdots \\ dC_{nm} \\ \vdots \\ dS_{nm} \\ \vdots \end{bmatrix} , \quad (6.5a)$$

where the pertinent partial derivatives have been expressed in (4.6). The above equation can be written in the following compact form:

$$dr = \left| \partial r/\partial(C's, S's) \right| \ d(C's, S's) \quad .$$

Upon applying the law of the variance-covariance propagation and upon considering (6.2), we obtain

$$\sigma_N^2 = \left| \partial r/\partial(C's, S's) \right| \sum_{C's,S's} \left| \partial r/\partial(C's, S's) \right|^T . \quad (6.5b)$$

In a similar fashion, we express

$$d\Delta g = \left| \ldots \partial \Delta g/\partial C_{n0} \ldots, \ \ldots \partial \Delta g/\partial C_{nm} \ldots, \ \ldots \partial \Delta g/\partial S_{nm} \ldots \right| \begin{bmatrix} \vdots \\ dC_{n0} \\ \vdots \\ dC_{nm} \\ \vdots \\ dS_{nm} \\ \vdots \end{bmatrix} , \quad (6.6$$

where the partial derivatives have been given in (5.27). In the compact form, this is written as

$$d\Delta g = \left| \partial \Delta g/\partial(C's, S's) \right| \ d(C's, S's) \quad ;$$

hence

$$\sigma_{\Delta g}^2 = \left| \partial \Delta g/\partial(C's, S's) \right| \sum_{C's, S's} \left| \partial \Delta g/\partial(C's, S's) \right|^T . \quad (6.6b)$$

The formulas (6.5b) and (6.6b) are used to generate a grid of a-posteriori standard errors in geoid undulations and in gravity anomalies. Such grids may serve to generate contour maps. However, these standard errors must be understood as associated with a specific mathematical model. In particular, they strictly correspond to an idealized earth which does not admit higher degree and order potential coefficients than those determined by the chosen degree and order truncation (all the other coefficients are assumed to be perfect zeros). The standard errors in the real world would be of course higher, but the present values are also useful. They can serve in indicating areas of a weak adjustment, in assessing the contribution of added data, in various comparisons, etc. Their real meaning, however, should be kept in mind.

7. GLOBAL ADJUSTMENT OF REAL SATELLITE ALTIMETRY AND GRAVITY ANOMALY DATA

7.1 Description of Data and Certain Adjustment Characteristics

The final, improved version of the program SAGG has been used, for the first time, in a combined adjustment of real satellite altimetry data and real gravity anomaly data. The satellite altimetry data consists of two basic sets of GEOS-3 altimetry observations. The first set includes data from one-hundred and twelve passes of the GEOS-3 satellite over a portion of the North Atlantic. The second set includes data from one-hundred and eleven passes over adjacent areas of the Indian and South Pacific Oceans. The gravity anomaly data is represented by mean anomalies from over 2200 $1^0 \times 1^0$ geographic blocks covering many continental as well as oceanic regions. We mention that a more detailed description of the GEOS-3 ground tracks as well as their graphic representation on a global scale may be found in the AFGL in-house paper, "Combination of Satellite Altimetry Data in the Short Arc Mode and Gravity Anomaly Data" (in press).

During the initial preprocessing of the GEOS-3 data, certain characteristics were identified for editing criteria. The first editing level automatically examined altimetry measurements for gross errors that occurred from data handling procedures, such as tape parity errors and measurement identification problems. The second editing procedure was intended to examine the altimeter measurements for continuity and to eliminate abrupt point to point changes. The final editing is based on a "three sigma" criteria when compared to a polynomial smoothing function.

The position a-priori standard errors of the state vectors have been assumed to be 24 m, 17 m, and 8 m in the in-track, cross-track, and (approximately) "up" directions, respectively. The velocity standard errors have been similarly assumed to be 0.02 m/sec , 0.02 m/sec , and 0.01 m/sec. The a-priori standard error for all the altimetry measurements has been assumed to be 1.0 meter. On the other hand, the a-priori standard error of each gravity anomaly has been read from a magnetic tape which also contained the gravity anomaly value itself and the positional information (ϕ, λ coordinates of the N - W corner of each gravity anomaly block).

The model that has served in the combined global adjustment is characterized by the presence of spherical harmonic potential coefficients up to and including the degree and order (14, 14). The option to include a-priori standard errors in the coefficients has been exercised in the following manner:

$$\left.\begin{aligned} \sigma_{C_{20}} &\simeq 1.1 \times 10^{-6} \qquad (\simeq 0.001\,|C_{20}|)\ , \\ \sigma_{C_{22}} &\simeq 1.6 \times 10^{-7} \qquad (\simeq 0.1\ \ |C_{22}|)\ , \\ \sigma_{S_{22}} &\simeq 1.0 \times 10^{-7} \qquad (\simeq 0.1\ \ |S_{22}|)\ . \end{aligned}\right\} \qquad (7.1)$$

The five "forbidden" potential coefficients have been held fixed at the zero value. The parameters of the normal field that have been adopted in this adjustment correspond to the mean earth ellipsoid as described in |Moritz, 1975|.

Throughout most of the present analysis, the quantity r_0, defined as

$$r_0 = kM/W_0 \quad , \tag{7.2}$$

has been considered constant. However, the combined adjustment described in this chapter is an exception to the rule. In this instance, r_0 is regarded as an adjustable parameter. Although (7.2) still represents the original value of r_0, where

$$W_0 = U_0$$

is assumed to hold as in (5.3), etc., this new parameter is allowed a certain freedom to adjust, depending on the assigned weight. The only changes in the adjustment formulas occur at the level of partial differentiation. In particular, $\partial r/\partial r_0$ must now be added to the row of the partial derivatives $\partial r/\partial(C's, S's)$, appearing in (4.7) as well as in equations (6.5) which deal with the variance-covariance propagation. Similarly, $\partial\Delta g/\partial r_0$ must be added to the row $\partial\Delta g/\partial(C's, S's)$, appearing in (5.28) as well as in equations (6.6). The location of such additional partial derivatives must correspond to the location of the new correction dr_0 among the familiar corrections to the potential coefficients (dC's, dS's).

In order to evaluate $\partial r/\partial r_0$, from (4.4) we observe that if r_0 becomes a parameter, the quantity $dr_0(1 + P_1 + \frac{1}{2}D)$ is added to the right-hand side of both (4.2) and (4.5a). It thus follows that

$$\partial r \,/\, \partial r_0 = (1 + P_1 + \tfrac{1}{2} D) \,/\, Q \quad . \tag{7.3}$$

On the other hand, upon specializing (2 - 179') of $\left|HM\right|$ to our needs, we deduce that the differential increment in Δg due to dW_0 is $(2/R)dW_0$. From

(7.2) we observe that in terms of dr_o, this increment is $-(2/R)(kM/r_o^2)dr_o$, or approximately $-(2/r_o)Cdr_o$. We thus deduce that

$$\partial \Delta g / \partial r_o \simeq -(2/r_o)\, C \quad . \tag{7.4}$$

The weights in all of the present real data reductions have been formulated according to (7.1), supplemented by

$$\sigma_{r_o} = 1 \text{ m} \quad .$$

Due to the scarcity of satellite altimetry data, a separate altimetry adjustment in the (14, 14) geopotential model failed (the system of normal equations was severely ill-conditioned). The gravity anomaly data has been adjusted alone and in combination with the altimetry data. However, the present gravity anomaly data also exibits gaps, although to a lesser extent than the satellite altimetry data; furthermore, in several areas, the a-priori standard error is extremely large. This is locally reflected by relatively large a-posteriori standard errors in geoid undulations and gravity anomalies, both in the separate gravity anomaly adjustment and in the combined adjustment.

7.2 Presentation of the Results

The inclusion of satellite altimetry data has significantly improved the external reliability of the gravity anomaly adjustment. This is due not only to the addition of data in areas where it had been previously lacking, but also to the fact that a completely different and independent source of data has been allowed to play its role in the determination of the earth's gravity field. As far as the internal precision is concerned, significant improvements have been noted, especially in the areas of the thus added data. This precision is represented by a-posteriori standard errors in geoid undulations and in gravity anomalies. As an example, the standard error in geoid undulations in the Indian Ocean region (around $\phi = -60^0$ and $\lambda = 120^0$) improved more than ten times (from 8.3 m to 0.82 m). The internal precision is of course invariably too optimistic with regard to the "real world" since it reflects the assumption that the geoid is exactly represented by the spherical harmonic expansion truncated at a given degree and order (any further coefficients as well as their standard errors are considered to be exactly zero). However, the a-posteriori standard errors have been of great assistance in depicting areas of insufficient or weak data, etc.

The combined adjustment in the (14, 14) geopotential model has yielded the values of geoid undulations (N), gravity anomalies (Δg), a-posteriori standard errors in geoid undulations (σ_N), and a-posteriori standard errors in gravity anomalies ($\sigma_{\Delta g}$), all computed in a $10^0 \times 10^0$ geographic grid. These values have been subsequently used for plotting contour maps. Figure 7.1 and Figure 7.2 show the contour maps for N (contour interval is 10 m) and Δg (contour interval is 10 mgal), respectively. The areas of

insufficient data on these maps have been blanked out. The "cut-off" decision for this purpose has been facilitated precisely by the contour maps of standard errors. A certain contour (in this case the 3 m contour) has been chosen from the plot associated with σ_N, beyond which the data have been deemed insufficient. The values of σ_N in the areas of weak geoid determination rise sharply and are three to ten fold larger than the values in all other parts of the globe.

The geoid contour map has been compared with the geoidal maps appearing in |Hadgigeorge and Trotter, 1976|. In the areas of comparison, the shape of the geoid exhibits remarkable similarities. In addition, the numerical values of geoid undulations themselves agree very satisfactorily. It is to be noted that the geoid presented in this reference had in turn been compared, along four different profiles, with a detailed gravimetric geoid (see Section 3.2 and 3.3 therein); the results of these comparisons were reported as excellent. The present data reduction is thus verified from independent sources.

In order to provide more insight into the interdependence of the degree and order of truncation and the detail in geoidal features, a new combined adjustment has been performed, this time in conjunction with a (13, 13) geopotential model. This has yielded a new set of values N, Δg, σ_N, and $\sigma_{\Delta g}$. The pertinent contour maps for N and Δg appear respectively in Figure 7.3 and Figure 7.4, in which the same areas have been blanked out as previously. An interesting comparison can be made between the Figures 7.1 and 7.3 featuring the geoid undulations, and between the Figures 7.2 and 7.4

featuring the gravity anomalies. These figures reveal a certain technical deficiency, in that the contour lines are not always pleasingly smooth. However, this lack of smoothness could be easily corrected if the predictions were made e.g. in a $5^0 \times 5^0$ geographic grid rather than in the coarse $10^0 \times 10^0$ grid.

When comparing the Figures 7.1 and 7.3, we notice many similarities, both global and local, between the geoid undulations in the (14, 14) and (13, 13) models. A reasonable amount of smoothing in Figure 7.3 versus Figure 7.1 may be observed over Japan and the surrounding region. In other areas, smoothing, although generally present, is not as easily observed; in several instances, the peaks in Figure 7.3 are actually higher than the corresponding features in Figure 7.1. However, the overall picture that emerges from this comparison is not significantly different from what can be expected when the truncation in two models differs by no more than one degree.

The smoothing in gravity anomalies in Figure 7.4 versus Figure 7.2 appears to be more pronounced than in the case with geoid undulations. The area where the smoothing is best observed includes again Japan, but extends further south and south-east. The peaks in the (13, 13) model are most of the time lower than their counterparts in the (14, 14) model. In several instances, this comparison offers some typical examples of smoothing.

The contours in standard errors, although available, are not presented here. The most useful in this respect has been the 3 m "cut-off" contour of σ_N in the (14, 14) model. As expected, consistently lower

standard errors have been noticed in the (13, 13) model when compared to the (14, 14) model, both in σ_N and $\sigma_{\Delta g}$. This is explained by the fact that fewer parameters are present in the former model than in the latter; in particular, all the parameters of degree 14 are considered to be perfect zeros in the (13, 13) model.

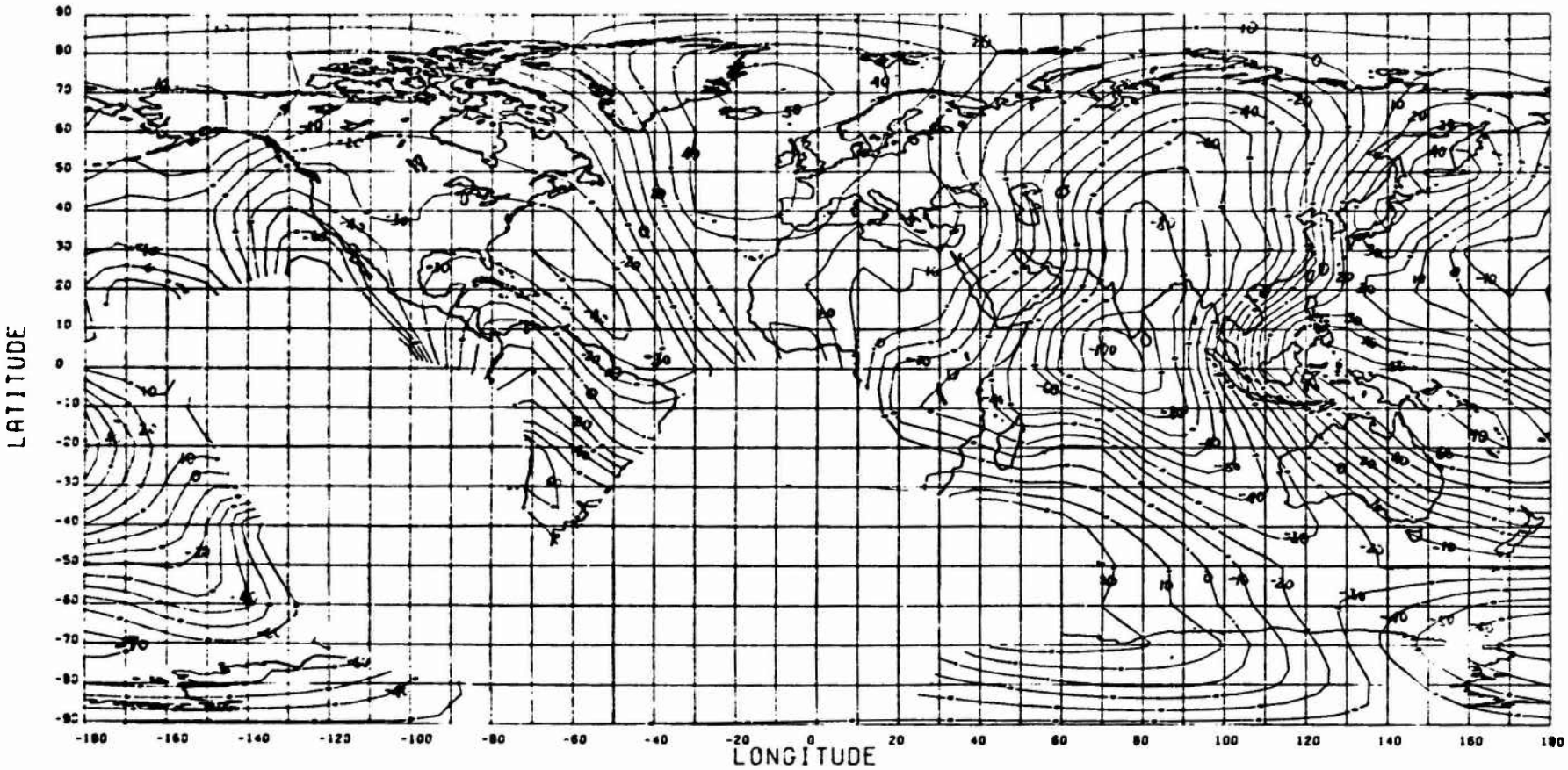


FIGURE 7.1

Ten meter isocontours of geoid undulations recovered from SAGG reduction of satellite altimetry and mean gravity anomalies in a spherical harmonic model truncated at degree and order (14, 14)

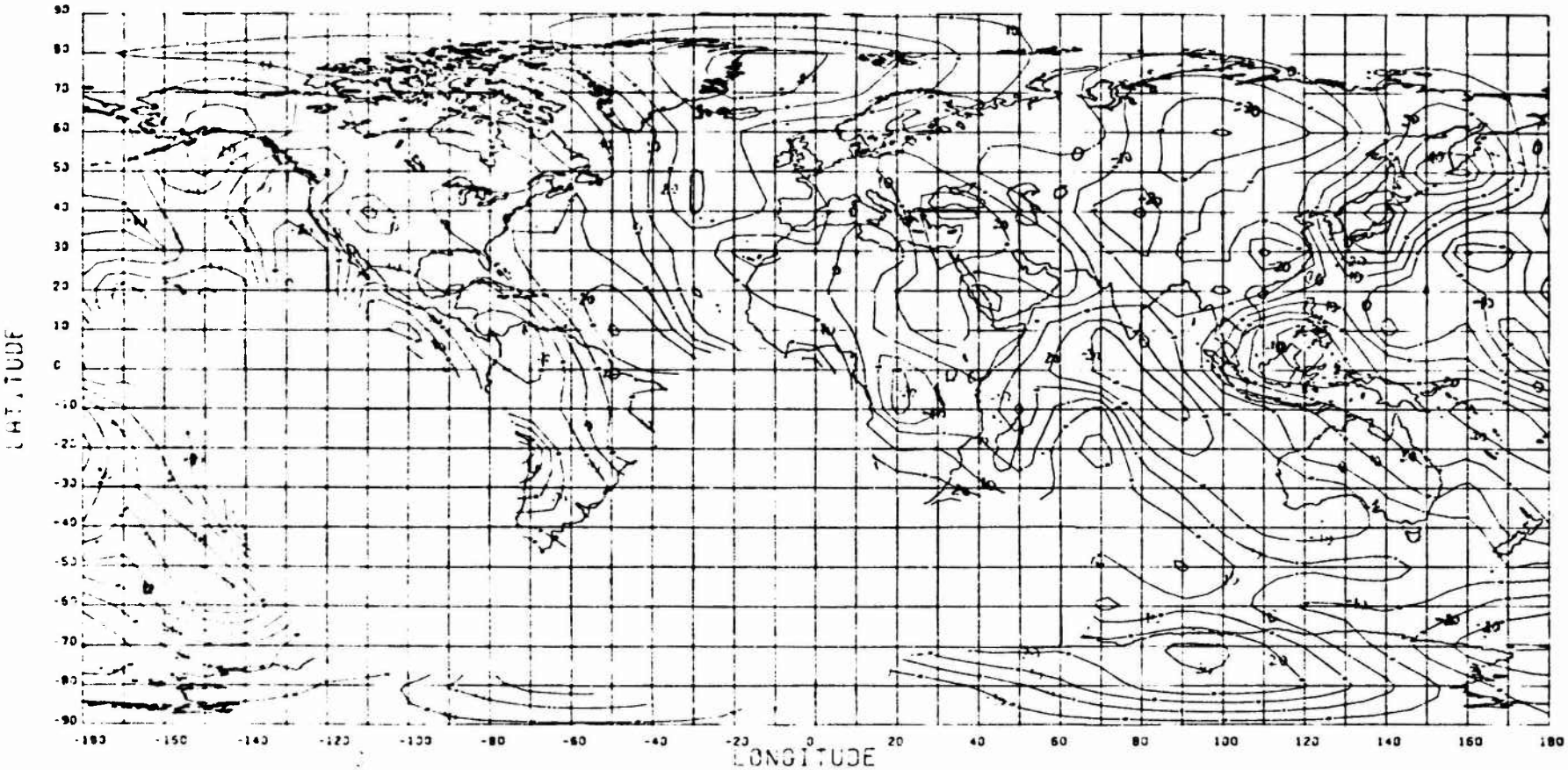


FIGURE 7.2

Ten milligal isocontours of gravity anomalies recovered from SAGG reduction of satellite altimetry and mean gravity anomalies in a spherical harmonic model truncated at degree and order (14, 14)

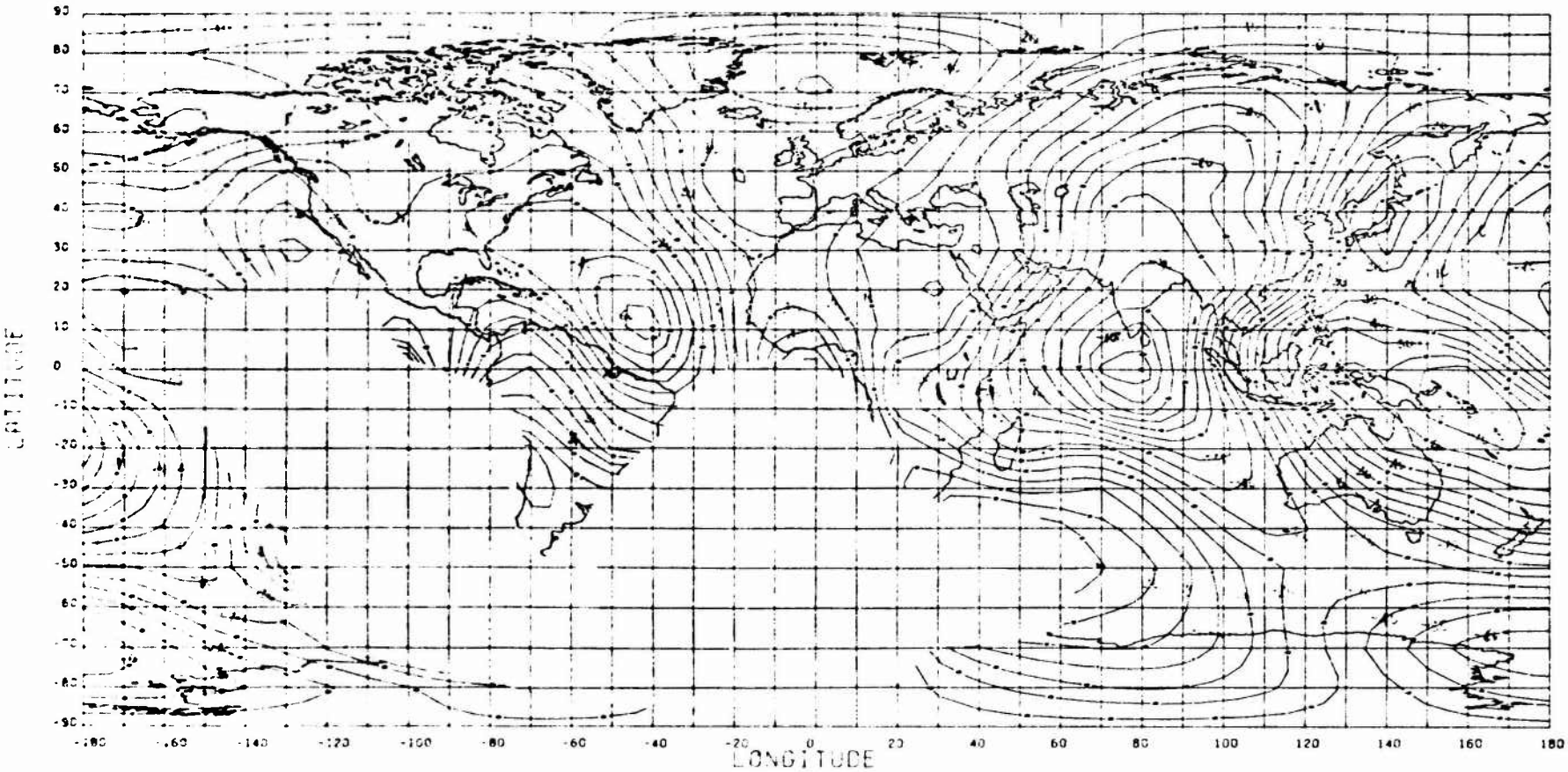


FIGURE 7.3

Ten meter isocontours of geoid undulations recovered from SAGG reduction of satellite altimetry and mean gravity anomalies in a spherical harmonic model truncated at degree and order (13, 13)

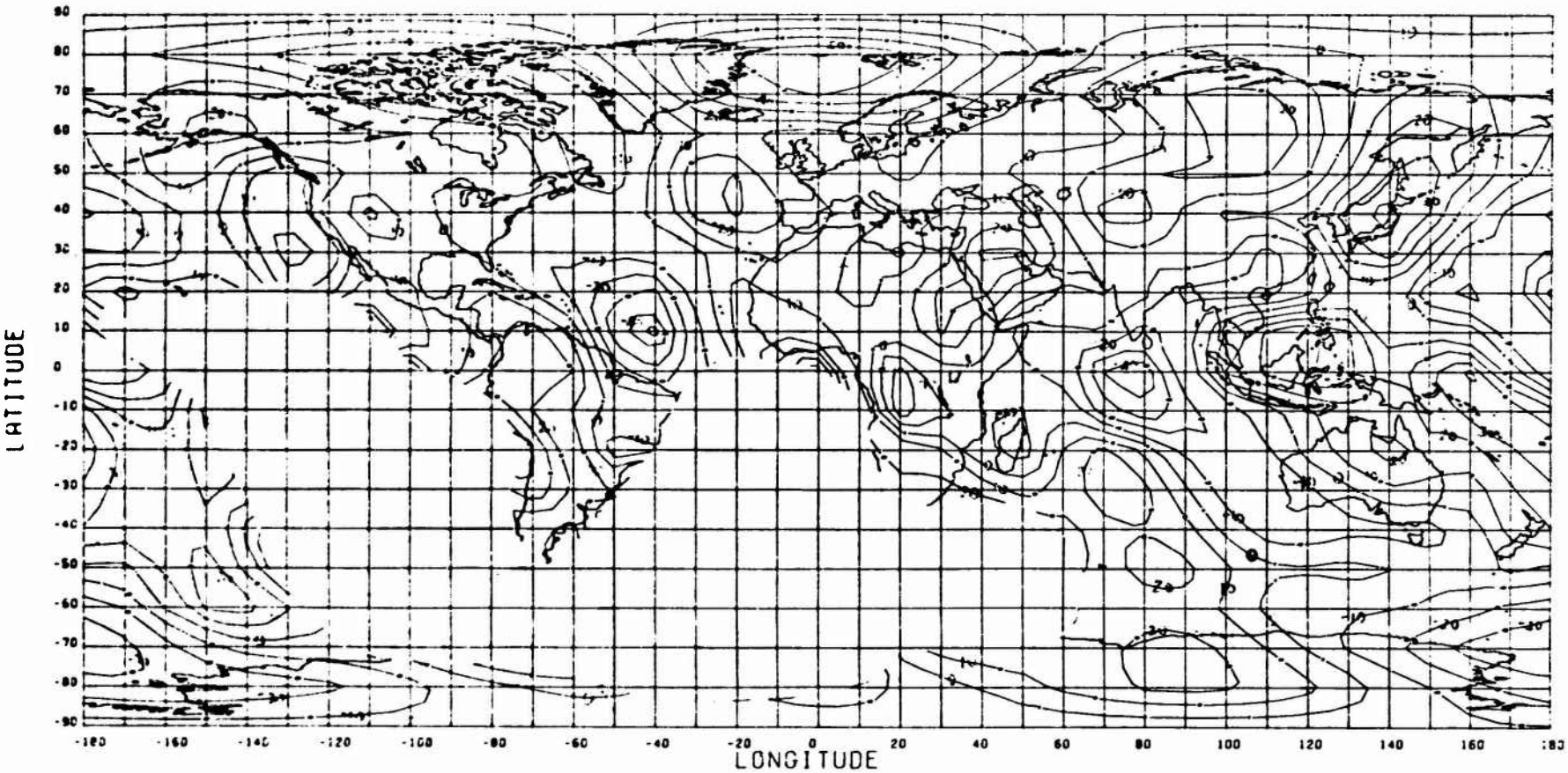


FIGURE 7.4

Ten milligal isocontours of gravity anomalies recovered from SAGG reduction of satellite altimetry and mean gravity anomalies in a spherical harmonic model truncated at degree and order (13, 13)

8. CONSIDERATIONS OF SPHERICAL VERSUS SPHEROIDAL HARMONICS

8.1 Practical Adjustment of Functions Expressed in Surface Spherical Harmonics

When considering an expansion of a function in a spherical harmonic series, a question arises whether there are other suitable coordinate systems (besides the spherical) in which a similar expansion could be carried out with better results. In particular, the spheroidal (or ellipsoidal) harmonics and other functions have been mentioned in this context. However, as related e.g. in [Heiskanen and Moritz, 1967], abbreviated here as [HM], page 39,

> "The whole matter is a question of mathematical convenience, since both spherical and ellipsoidal harmonics may be used for any attracting body, regardless of its form."

It follows that the use of spherical harmonics, spheroidal harmonics, or harmonic functions in yet other systems should yield the same theoretical results for the potential and related quantities whenever the pertinent series converge. The emphasis is thus shifted toward convergence problems and regions of convergence. To this effect, we may quote Needham [1970], page 46:

> "Current literature contains many examples of the use of spherical harmonic expansions of the potential field of the earth for the determinations of geoidal undulations, deflections of the vertical, and gravity anomalies. In these applications mention is usually made of convergence problems in a series ... when the point of interest is inside a sphere which just includes all attracting masses..."

However, before we start dealing with the convergence of spherical harmonic series and ponder the possibility of replacing them with spheroidal harmonic series, some practical aspects of spherical harmonic expansions should be understood, sincethe theoretical findings pertinent to the infinite series in the "true" spherical harmonic coefficients may not be entirely relevant when one starts dealing with the least squares (L.S.) adjustment of these coefficients in practice. It is in this spirit that we shall first discuss theoretical formulas

depicting quantities (e.g. gravity anomalies, geoid undulations) expressed in terms of surface spherical harmonics; for the moment, the convergence problems will be disregarded. We shall then gradually pass into the realm of real observations and the practical determination of the spherical harmonic potential coefficients.

An expansion of a function Y in surface spherical harmonics can be written in a fashion similar to [HM], Sections 1-13, 1-14:

$$Y \equiv Y(\bar{\phi}, \lambda) = \sum_{n=0}^{\infty} Y_n = \sum_{n=0}^{\infty} \sum_{m=0}^{n} (\bar{c}_{nm} \bar{R}_{nm} + \bar{d}_{nm} \bar{S}_{nm}) , \quad (8.1a)$$

$$\begin{Bmatrix} \bar{c}_{nm} \\ \bar{d}_{nm} \end{Bmatrix} = \frac{1}{4\pi} \iint_{\sigma} Y \begin{Bmatrix} \bar{R}_{nm} \\ \bar{S}_{nm} \end{Bmatrix} d\sigma , \quad (8.1b)$$

where

$$\begin{aligned} \bar{R}_{nm} \equiv \bar{R}_{nm}(\bar{\phi},\lambda) &= \bar{P}_{nm}(\sin\bar{\phi}) \cos m\lambda , \\ \bar{S}_{nm} \equiv \bar{S}_{nm}(\bar{\phi},\lambda) &= \bar{P}_{nm}(\sin\bar{\phi}) \sin m\lambda , \end{aligned} \quad (8.2)$$

and where $\bar{P}_{nm}$ denotes the associated Legendre functions in a normalized form. Here and in the following formulas, the equation "a" is used to compute the values of Y from known coefficients ($\bar{c}_{nm}$ and $\bar{d}_{nm}$), while the equation "b" gives the coefficients from known values of Y at every point on the unit sphere denoted by the symbol σ . For the zero-degree harmonic, we have

$$Y_o \equiv \bar{c}_{oo} = (1/4\pi) \iint_{\sigma} Y \, d\sigma .$$

Specializing the above equations to gravity anomalies, we obtain

$$\Delta g = \Delta g_o + G \sum_{n=2}^{\infty} (n - 1) \sum_{m=0}^{n} (\overline{\Delta C}_{nm} \bar{R}_{nm} + \overline{\Delta S}_{nm} \bar{S}_{nm}), \quad (8.3a)$$

where $G \simeq 979.8$ gal is the average value of gravity and $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$ are the normalized coefficients of the disturbing potential referred to a given ellipsoid; $\overline{\Delta S}_{nm}$ should not be confused in any way with $\overline{S}_{nm}$ described in (8.2). Equation (8.3a), without the term Δg_0, is essentially equation (6) of [Rapp, 1970]. According to [HM], equations (2-185), (2-186), this term can be expressed as

$$\Delta g_0 \equiv (1/4\pi) \iint_\sigma \Delta g \, d\sigma = - (1/R^2) \, k\delta M + (2/R) \, \delta W \quad ,$$

where $k\delta M$ is the correction to kM' of the reference ellipsoid in order to obtain kM (gravitational constant times mass) of the earth, and δW is the correction to the potential on the reference ellipsoid in order to obtain the numerical value of the potential on the geoid. For $n \geq 2$, in agreement with (8.1b) we have

$$\begin{Bmatrix} \overline{\Delta C}_{nm} \\ \overline{\Delta S}_{nm} \end{Bmatrix} = \frac{1}{4\pi \, G(n-1)} \iint_\sigma \Delta g \begin{Bmatrix} \overline{R}_{nm} \\ \overline{S}_{nm} \end{Bmatrix} d\sigma \, . \qquad (8.3b)$$

This equation also appeared in [Rapp, 1970] as equation (8). Considering the last two equations, we conclude that the computation of $k\delta M$ or δW (not both), and of $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$ would be possible if, in theory, the gravity anomalies were known everywhere on the earth.

If the center of the reference ellipsoid is located at the earth's center of mass, we can similarly write for geoid undulations:

$$N = N_0 + R \sum_{n=2}^{\infty} \sum_{m=0}^{n} (\overline{\Delta C}_{nm} \, \overline{R}_{nm} + \overline{\Delta S}_{nm} \, \overline{S}_{nm}) \quad , \qquad (8.4a)$$

where $R \simeq 6371$ km is the earth's mean radius. Without the term N_0, this equation corresponds to equation (20) in [Rapp, 1972], after the spherical approximation has been introduced. We also have

$$N_0 \equiv (1/4\pi) \iint_\sigma N \, d\sigma = k\delta M/(RG) - \delta W/G \quad ,$$

where the right-hand side appears in [HM], equation (2-182). From (8.1b), for $n \geq 2$ it follows that

$$\begin{Bmatrix} \overline{\Delta C}_{nm} \\ \overline{\Delta S}_{nm} \end{Bmatrix} = \frac{1}{4\pi R} \iint_{\sigma} N \begin{Bmatrix} \overline{R}_{nm} \\ \overline{S}_{nm} \end{Bmatrix} d\sigma \quad . \tag{8.4b}$$

Here again, the computation of $k\delta M$ or δW (not both), and of $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$ would be possible if, in theory, the geoid undulations were known everywhere on the earth.

Before discarding the corrections $k\delta M$ and δW, i.e. before assuming -- as elsewhere in this analysis -- that the reference ellipsoid has the same mass as the actual earth and the same potential as the geoid, we mention that both $k\delta M$ and δW could be computed if both Δg_0 and N_0 were known; one could then proceed according to (2-187a), (2-187b) of [HM]. An independent determination of N_0 could be made with the aid of at least one distance and the astronomical latitude and longitude at the endpoints, as described in Section 2-19 of [HM]. However, the important point we want to emphasize in this context is the global distribution of gravity data both for the determination of kM and W^0 (notation used in [HM] for the potential on the geoid) and of $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$. We use two quotations from [HM] to this effect; the first appears on page 107:

> "Another significant fact is that since the gravity anomalies on the whole earth are needed ..., the constants kM and W^0 cannot be determined unless gravity g is known all over the earth. This again reflects the general principle of the gravimetric method — namely, that it is always required that g be known at every point of the earth's surface."

As for the coefficients $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$, on page 109 we read

> "Hence the determination of the spherical-harmonic coefficients of the earth's potential can be described as follows. We expand the gravity anomalies Δg, which must be given all over the world, as a series of spherical harmonics..."

Nothing essential will be lost if we assume

$$k\delta M \equiv 0 \quad , \qquad \delta W \equiv 0 \quad ,$$

and if we limit ourselves to gravity anomalies in the upcoming demonstrations dealing with practical aspects of the determination of the earth's gravity field. We thus have

$$\Delta g_0 \equiv 0 \ . \tag{8.5}$$

The globe may now be visualized subdivided into Q small blocks of equal area whose size is represented by the central angle $\Delta\theta$. The size is assumed sufficiently small so that the integration may be replaced by a summation. Each block has associated with it one value of Δg (a certain mean anomaly). From (8.3b), the numerical integration thus yields

$$\begin{Bmatrix} \overline{\Delta C_{nm}} \\ \overline{\Delta S_{nm}} \end{Bmatrix} = \frac{1}{G(n-1)} \left[\frac{1}{Q} \sum_{i=1}^{Q} \Delta g_i \begin{Bmatrix} \bar{R}_{nm}(i) \\ \bar{S}_{nm}(i) \end{Bmatrix} \right] , \tag{8.6b}$$

where, in this "b" - type equation,

$$\bar{R}_{nm}(i) \equiv \bar{R}_{nm}(\bar{\phi}_i, \lambda_i) \ ,$$
$$\bar{S}_{nm}(i) \equiv \bar{S}_{nm}(\bar{\phi}_i, \lambda_i) \ ,$$

and where $\bar{\phi}_i$, λ_i refer to the center of the i th $\Delta\theta \times \Delta\theta$ block. The brackets on the right-hand side contain the global means of $\Delta g\bar{R}_{nm}$ and $\Delta g\bar{S}_{nm}$. In a first example to follow, the coefficients will be determined up to and including the degree and order (k, k), their total number being equal to the number of gravity anomalies (their total number is $(k+1)^2$ if no coefficient is excluded). The value of k may be approximated as

$$k \simeq 180^\circ / \Delta\theta^\circ$$

(see e.g. [Rapp, 1970], equation 7). Higher degree and order coefficients are set to zero. This is consistent with the system of Q linear equations (8.3a)

considered with (8.5), written now as

$$\Delta g_i = G \sum_{n=2}^{k} (n-1) \sum_{m=0}^{n} \left[\overline{\Delta C}_{nm} \overline{R}_{nm}(i) + \overline{\Delta S}_{nm} \overline{S}_{nm}(i) \right], \quad (8.6a)$$

which could not yield a unique solution for a set of coefficients larger than the one symbolized by (k, k).

From the coefficients computed by (8.6b), equation (8.6a) could be used to compute Δg at any point. The shortest half wavelength of geoidal features that could still be expressed is roughly $\Delta\theta$. The features of shorter wavelengths than $2\Delta\theta$ would be suppressed, resulting in "true" smoothed functions. If (8.6a) is used for any of the points $\overline{\phi}_i$, λ_i, the original Δg_i entering (8.6b) should be recovered since no adjustment would take place in the first example involving a consistent system (the only deviations could be due to the numerical integration). This will be illustrated in the next paragraph through a L.S. adjustment, where Δg_i will be considered as observations (denoted Δg_i^b) of equal precision. The observation equations in standard matrix notation can be represented by

$$V = AX - \Delta g^b \quad ,$$

where V is the residual vector, A is the "design" matrix, X is the parameter vector containing $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$, and Δg^b is the vector containing Δg_i^b. If we solve this system by the regular L.S. method, namely

$$X = (A^T A)^{-1} A^T \Delta g^b \quad , \qquad (8.7)$$

we have in fact solved the original system

$$X = A^{-1} \Delta g^b \quad ,$$

because A is a square, nonsingular matrix. But this means that V = 0 and no adjustment of gravity anomalies takes place.

Our attention will be now focused on an analytical solution of the first example (it involves a specific consistent system of Q observations) through a L.S. adjustment. Some features of this example will prove useful also later, in a second example, where the number of parameters will be smaller than the number of observations. In either case, the observations are assumed to be generated by (8.6a). Our immediate task is to show that the L.S. solution indeed recovers the original (k, k) set of potential coefficients (8.6b). The observations are

$$\Delta g_i^b = G \sum_{n=2}^{k} (n-1) \sum_{m=0}^{n} \left[\overline{\Delta C}_{nm} \overline{R}_{nm}(i) + \overline{\Delta S}_{nm} \overline{S}_{nm}(i) \right] , \quad 1 \leq i \leq Q \quad . (8.8)$$

An i th observation equation is written as

$$v_i = a_i X - \Delta g_i^b ,$$

or

$$v_i = \left[\ldots G(n-1)\, \overline{R}_{nm}(i) \ldots ; \ldots G(n-1)\, \overline{S}_{nm}(i) \ldots \right] \begin{bmatrix} \vdots \\ \overline{\Delta C}_{nm} \\ \vdots \\ \hline \vdots \\ \overline{\Delta S}_{nm} \\ \vdots \end{bmatrix} - \Delta g_i^b \quad . \qquad (8.9)$$

The row vector (a_i) on the right-hand side of (8.9) has as many elements as there are parameters; in this case it is Q elements corresponding to $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$ complete through the degree and order (k, k). The number of observation equations is always Q. The matrix A is formed by joining together all of a_i; it thus has Q rows.

In computing the matrix of normal equations -- denoted as (A^T A) in (8.7) -- use will be made of orthogonality relations that appear on pages 29 and 31 of [HM]; in particular,

$$\underbrace{\iint_\sigma \bar{R}_{nm}\bar{R}_{sr}d\sigma = 0\ ,\quad \iint_\sigma \bar{S}_{nm}\bar{S}_{sr}d\sigma = 0}_{\text{if } s \neq n \text{ or } r \neq m \text{ or both}}\ ,\quad \underbrace{\iint_\sigma \bar{R}_{nm}\bar{S}_{sr}d\sigma = 0}_{\text{in any case}}\ ,$$

$$\iint_\sigma \bar{R}^2_{nm}d\sigma = 4\pi\ ,\quad \iint_\sigma \bar{S}^2_{nm}d\sigma = 4\pi\ .$$

In connection with the present numerical integration we thus have

$$\underbrace{\sum_{i=1}^{Q} \bar{R}_{nm}(i)\bar{R}_{sr}(i) = 0,\quad \sum_{i=1}^{Q} \bar{S}_{nm}(i)\bar{S}_{sr}(i) = 0}_{\text{if } s \neq n \text{ or } r \neq m \text{ or both}},\quad \underbrace{\sum_{i=1}^{Q} \bar{R}_{nm}(i)\bar{S}_{sr}(i) = 0,}_{\text{in any case}}$$

$$\sum_{i=1}^{Q} \bar{R}^2_{nm}(i) = Q\ ,\qquad \sum_{i=1}^{Q} \bar{S}^2_{nm}(i) = Q\ .$$

With the help of these orthogonality relations, the matrix of normal equations is readily constructed from (8.9) as

$$A^T A = \begin{bmatrix} \ddots & & \\ & G^2\,(n-1)^2\,Q & 0 \\ & & \ddots \\ \hline & & \ddots \\ & 0 & G^2\,(n-1)^2\,Q \\ & & & \ddots \end{bmatrix}\ . \qquad (8.10)$$

Each of the two diagonal blocks would have exactly the same form -- but smaller dimensions -- if the number of parameters decreased. This follows from the form of the row vector (a_i) in (8.9); the number of columns in A would be smaller, but each remaining column would participate in exactly

the same multiplications and summations that occurred in forming (8.10).

The matrix of normal equations in (8.10) is diagonal; we thus have

$$(A^T A)^{-1} = \text{diag}\left\{\ldots\ 1/\left[G^2(n-1)^2\ Q\right]\ldots\ \vdots\ \ldots\ 1/\left[G^2(n-1)^2\ Q\right]\ \ldots\right\}. \tag{8.11}$$

The column vector $A^T\ \Delta g^b$ is formed in a straightforward manner as

$$A^T\ \Delta g^b = \begin{bmatrix} \vdots \\ G(n-1) \sum_{i=1}^{Q} \Delta g_i^b\ \bar{R}_{nm}(i) \\ \vdots \\ \hline \vdots \\ G(n-1) \sum_{i=1}^{Q} \Delta g_i^b\ \bar{S}_{nm}(i) \\ \vdots \end{bmatrix} . \tag{8.12}$$

In analogy to our previous statement, we notice that if the number of parameters decreased, each of the two parts of the column vector in (8.12) would be shortened, but the remaining elements would be exactly the same.

Carrying out the matrix multiplication between (8.11) and (8.12) as indicated in (8.7), we obtain

$$\begin{bmatrix} \vdots \\ \overline{\Delta C}_{nm} \\ \vdots \\ \hline \vdots \\ \overline{\Delta S}_{nm} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \left[1/G\ (n-1)\right](1/Q) \sum_{i=1}^{Q} \Delta g_i^b\ \bar{R}_{nm}(i) \\ \vdots \\ \hline \vdots \\ \left[1/G\ (n-1)\right](1/Q) \sum_{i=1}^{Q} \Delta g_i^b\ \bar{S}_{nm}(i) \\ \vdots \end{bmatrix} , \tag{8.13}$$

which is precisely (8.6b); the superscript b in (8.13) merely indicates that Δg_i were treated as adjustable quantities, although we knew beforehand that no adjustment per se would take place, in view of the discussion that followed (8.7). We have thus far demonstrated that as long as the equal area blocks are sufficiently small so that the orthogonality relations remain valid, there is a perfect agreement between the equations of the type "a" and "b". On the other hand, if such relations were not fulfilled, the normal equation matrix in (8.10) would change while the column vector in (8.12) would remain unchanged. Consequently, the solution would in general change and (8.6b) would not be recovered. We further notice that in the case of a square nonsingular matrix A such as treated presently, weighting of observations (the weight matrix is denoted as P) cannot affect the outcome of the adjustment, since

$$(A^T P A)^{-1} A^T P \Delta g^b = A^{-1} \Delta g^b ;$$

but this is the same result as the one obtained in (8.7).

We can now pass to the more practical second example in which the number of parameters is smaller than the number of observations and, consequently, a L.S. filtering takes place. The set of observations of equal precision is assumed to be exactly the same as in the previous case, i.e., it is generated in (8.8) via the same "true" coefficients $\overline{\Delta C}_{nm}$, $\overline{\Delta S}_{nm}$. However, the adjustable parameters are now represented by the coefficients through degree and order (k', k'), where

$$k' < k \quad .$$

In the next paragraph, we shall demonstrate that although the number of parameters is smaller than previously, the adjusted coefficients have precisely the same "true" values. The other coefficients in the new model are suppressed (set to zero). Under this proposition, the adjustment model yields a "true" smoothed function represented by the adjusted quantities Δg_i^a:

$$\Delta g_i^a \equiv \Delta g_i^b + v_i = G \sum_{n=2}^{k'} (n-1) \sum_{m=0}^{n} \left[\overline{\Delta C}_{nm}\bar{R}_{nm}(i) + \overline{\Delta S}_{nm}\bar{S}_{nm}(i)\right]. \quad (8.14)$$

Since in general $v_i \neq 0$, the original "true" values Δg_i^b do not coincide with the adjusted values.

Using the adjustment model (8.14), we form the observation equations of the same form as (8.9), except that the number of elements in a_i -- and thus the number of columns in A -- corresponds to the new, smaller number of parameters. We can now apply the remark made after (8.10) and form the matrix $(A^T A)^{-1}$ exactly as in (8.11) but of dimensions compatible with the degree and order (k', k'). The same reasoning holds with respect to (8.12) and, finally, with respect to (8.13) as well. We thus confirm that all the potential coefficients present in the new model are adjusted to take on their "true" values, the parameters beyond the degree and order (k', k') having been suppressed. Comparing (8.8) with (8.14), the residuals are seen to be

$$v_i = -G \sum_{n=k'+1}^{k} (n-1) \sum_{m=0}^{n} \left[\overline{\Delta C}_{nm}\bar{R}_{nm}(i) + \overline{\Delta S}_{nm}\bar{S}_{nm}(i)\right], \quad (8.15)$$

i.e., the contribution of higher degree and order coefficients than (k', k') has been filtered out of Δg_i^b by the adjustment. The word "true" has been written in quotes to indicate the underlying assumption that Δg_i^b are

considered as true values of gravity anomalies. As we shall see shortly, such true values do not automatically yield the true values of resolved parameters and thereby a true smoothed function. We may conclude the second example by observing that with the given ideal distribution of Δg_i^b, the L.S. procedure with equal weights preserves the physical meaning of all the recovered potential coefficients, no matter their number. This amounts to saying that the truncation alone does not destroy the physical meaning of the potential coefficients and does not reduce them to some regression parameters whose role would be merely to provide the best fit of a function to given observations.

Although it is not needed in the present discussion, the past two examples lead us to consider a situation that may result in useful simplications in some practical cases. We have seen that under certain conditions, a set of parameters may take on the same adjusted values whether it stands alone or is complemented by another set of parameters. In particular, upon assuming an idealized distribution of input (observed) values, etc., the common part of the adjusted potential coefficients $\overline{\Delta C}_{nm}$ and $\overline{\Delta S}_{nm}$ has been shown to be the same whether the degree and order (k', k') or (k, k) are considered, where k' < k. In such a situation, the additional coefficients (i.e., the second set of parameters) beyond the set symbolized by (k', k') could be adjusted separately, based on the residuals supplied by the adjustment of the first set.

Let $(\Delta g)_I$ and $(\Delta g)_{II}$ represent the adjusted values associated respectively with the first set of parameters (it corresponds to the degree and order k', k') and the second, complementary set of parameters (it consists

of the coefficients beyond the degree and order k', k', up to and including the degree and order k" such that k" $\leq$ k). Let the residuals in the simultaneous adjustment of both sets be represented by v', and the residuals resulting from an adjustment of the first set alone be represented by v. The observation equations for the simultaneous adjustment are written symbolically as

$$v' = (\Delta g)_I + (\Delta g)_{II} - \Delta g^b \quad , \tag{8.16}$$

while the observation equations associated with the first set are

$$v = (\Delta g)_I - \Delta g^b \quad . \tag{8.17}$$

Since by hypothesis $(\Delta g)_I$ is the same in (8.16) and (8.17), we can substitute the latter into the former, which yields

$$v' = (\Delta g)_{II} + v \quad ; \tag{8.18}$$

but this indicates that the residuals from the first adjustment have the role of minus new observations. In this way, we may resolve two smaller systems (8.17) and (8.18) instead of one large system (8.16). If, in practice, the distribution of observations is reasonably good, etc., the hypothesis of unchanged $(\Delta g)_I$ may be very beneficial from the computation point of view, and it may yield acceptable results at the same time.

The above approach will be now illustrated by another (third) example. Specializing (8.9) to our need, we write

$$v'_i = a'_i X' + v_i \quad , \qquad 1 \leq i \leq Q \quad , \tag{8.19}$$

where the primes are associated with the second set of potential coefficients

beyond the degree and order (k', k'). The L.S. solution of the first set within (k', k') has been carried out in the second example, from which it also follows (see 8.14) that

$$v_i = G \sum_{n=2}^{k'} (n-1) \sum_{m=0}^{n} \left[\overline{\Delta C}_{nm} \bar{R}_{nm}(i) + \overline{\Delta S}_{nm} \bar{S}_{nm}(i) \right] - \Delta g_i^b \quad . \quad (8.20a)$$

Further, we denote

$$a'_i = \left[\ldots G(n'-1)\bar{R}_{n'm'}(i) \ldots ; \quad \ldots G(n'-1)\bar{S}_{n'm'}(i) \ldots \right] , \quad (8.20b)$$

$$X' = \begin{bmatrix} \vdots \\ \overline{\Delta C}_{n'm'} \\ \vdots \\ \overline{\Delta S}_{n'm'} \\ \vdots \end{bmatrix} ; \quad (8.20c)$$

clearly,

$$n' > n \quad . \quad (8.21)$$

The matrix of normal equations is found exactly in the manner that led us to (8.10); the dimensions now correspond to the second set of parameters X' (see also the short discussion that followed 8.10). Accordingly, the inverse of this matrix is

$$(A'^T A')^{-1} = \text{diag} \{ \ldots 1/[G^2(n'-1)^2 Q] \ldots \vdots \ldots 1/[G^2(n'-1)^2 Q] \ldots \} . \quad (8.22)$$

The column vector representing the right-hand side of the normal equations is $-A'^T[v]$, the last symbol indicating a vector composed of the quantities v_i. For a given n, m in (8.20a), we have for n', m' associated with any column of A':

$$\sum_{i=1}^{Q} G^2(n'-1)(n-1) \, \overline{\Delta C}_{nm} \, \bar{R}_{n'm'}(i) \, \bar{R}_{nm}(i) = 0 \quad .$$

Such a sum of products is typical in forming $-A'^{T}[v]$ where, for the moment, only the first part of (8.20a) is taken into account, and it is a consequence of the orthogonality relations considered together with (8.21). Similar expressions would be of course obtained also for the products $\bar{R}_{n'm'}\ \bar{S}_{nm}$, etc., hence

$$-A'^{T}[v] = A'^{T}\Delta g^{b} \equiv \begin{bmatrix} \vdots \\ G(n'-1)\sum_{i=1}^{Q}\Delta g_i^b\,\bar{R}_{n'm'}(i) \\ \vdots \\ \hline \vdots \\ G(n'-1)\sum_{i=1}^{Q}\Delta g_i^b\,\bar{S}_{n'm'}(i) \\ \vdots \end{bmatrix} . \qquad (8.23)$$

Finally, the L.S. solution for the coefficients is obtained upon post-multiplying the matrix in (8.22) by the column vector in (8.23). The result has the same form as (8.13), except that all of n, m are replaced by n',m'. That is to say that the second set of parameters has recovered the "true" values as expected. Having illustrated the possibilibity that under certain ideal conditions the potential coefficients may be adjusted in parts, we shall now return to the main stream of our discussion.

In a real situation, the size of the blocks is in general larger than is admissible for the orthogonality relations. In addition, some blocks may be void of observations and, most often, the observational precision varies widely. In order to make a direct comparison with the second example, we first idealize this real situation by assuming the same blocks as previously and also by assuming that no observation is missing. It is clear that due to the

weight matrix such that

$$P \neq I ,$$

the L.S. adjustment will not yield, in general, the same numerical results as with the unit weight matrix which resulted in a diagonal matrix of normal equations, although the solution will be unbiassed assuming the unbiassed observations (see e.g. [Blaha, 1976], Section 3.2 for a general discussion). Subsequently, this property of unbiassedness will be destroyed due to the blocks being too large, missing observations, etc. The potential coefficients thus lose, at least partially, their true physical meaning and become to a certain extent regression parameters.

8.2 Suitability of Spherical Harmonics in Current Adjustment Procedures

The formulas expressing gravity anomalies, geoid undulations, etc., in surface spherical harmonics can be obtained from their counterparts in space where the series are assumed to converge. Perhaps the best illustration of this fact may be gathered from equations (2-155) and (2-155') of [HM]. In terms of solid spherical harmonics, we have

$$r\Delta g = \sum_{n=0}^{\infty} (n-1)(R/r)^{n+1} T_n (\bar{\phi}, \lambda) \quad ,$$

where Δg is the gravity anomaly at a point outside the earth, r, $\bar{\phi}$, λ are the three spherical coordinates of this point, R is the earth's mean radius, and T_n is the surface spherical harmonic of degree n; each T_n is a function of the earth's potential coefficients. This series is convergent, which means that after a certain term, the contribution of further terms is less and less significant. At the surface of the earth, r becomes smaller and, consequently, the above infinite series may become divergent. If we further approximate r on the earth's surface by R (the error introduced is considered negligible in this discussion), we obtain

$$\Delta g = (1/R) \sum_{n=0}^{\infty} (n-1)\, T_n (\bar{\phi}, \lambda) \quad .$$

This form represents the expansion of surface gravity anomalies in surface spherical harmonics and, as indicated, it may be divergent.

When discussing the convergence problems associated with the series just mentioned, the appeal must be made to the basic function that has served in deriving such series, i.e., to the potential. On page 60 of [HM], the

following discussion is found regarding the expansion of the gravitational potential (V) in terms of spherical harmonics:

> "For an arbitrary body, the expansion of V in spherical harmonics can be shown to converge always outside the smallest sphere $r = r_0$ that completely encloses the body ... If the earth were a homogeneous ellipsoid of about the same dimensions, then the series for V would indeed still converge at the surface of the earth. Owing to the mass irregularities, however, the series of the actual potential V of the earth must be considered divergent at the surface of the earth..."

These remarks concern the potential coefficients as infinite in number or at least including a very high degree and order. The problem of convergence changes with the truncation of the original infinite series. This fact is expressed by Hotine [1969], page 173, as follows:

> "In particular, it may be that the series at points on the topographic surface, although divergent, can be truncated at a certain number of terms to give a better answer than a formally convergent series would give for the same number of terms."

The convergence problems are in general alleviated when considering the spheroidal harmonics. A sphere of convergence is replaced by a spheroid of convergence as described by Hotine [1969], page 194:

> "In the case of the actual Earth, it is possible to choose a coordinate spheroid which just encloses all the matter and is generally much nearer to the topographic surface than any sphere that also encloses all the matter. Accordingly, we can say that the expression of the potential of the Earth in spheroidal harmonics can be made certainly convergent much nearer to the topographic surface than the potential expressed in spherical harmonics."

A theoretical remark to be made in this context is that gravity anomalies or geoid undulations (measured via satellite altimetry) refer to the geoid which is not outside the spheroid just enclosing all the masses, and the series must therefore be again considered formally divergent. However, it appears that

the truncation could be made at a higher order and degree than in the case with spherical harnomics before any problems connected with the convergence would be felt.

When considering the current practical applications at AFGL, we observe that the possible beneficial effect stemming from an eventual usage of spheroidal harmonics could hardly be noticed. The reason for this statement is that the truncation of the series has been usually made at relatively low degree and order; in particular, some recent real data reductions have been made with a (14, 14) truncated series. Beyond this, other representations of the potential have been considered, superimposed on the spherical harmonics model. If somewhat higher degree and order harmonics were used in such applications, they would probably not go much beyond the (20, 20) truncation. In any event, considering the practical needs with respect to the degree and order truncation, we do not have to be concerned with the convergence problems inherent in the spherical harmonic expansion. Needham [1970] states to this effect (pages 46, 47):

> "The earth does not have [the] property of homogeneity, however, and the series must be considered formally divergent at or near the terrestrial surface...This defect is more theoretical than actual in the usual applications since the series are truncated to yield smoothed approximations of the true functions."

If, for some reason, we decided to replace the spherical harmonic expansion of the potential by the spheroidal harmonic expansion, the increase in computing effort during actual data reductions would be formidable indeed. Rapp [1972] has studied the implications of such a task. As far as the gravity anomalies are concerned, the observation equations in terms of the

spheroidal harmonic potential coefficients would be formed according to his equations (28) and (27), which correspond respectively to (29.67) and (29.61) of [Hotine, 1969]. These two equations appear below in the form that is essentially a reproduction of the mentioned equations (28) and (27), followed by the identifications appearing on pages 18 and 19 of [Rapp, 1972]:

$$
\begin{aligned}
g_A\nu \;=\; & -\sum_{n=0}^{\infty}\sum_{m=0}^{n} (n-1)\, T_{nm} \\
& +\sum_{n=0}^{\infty}\sum_{m=0}^{n} \sin^2\alpha \cos^2 u\, T_{nm} \\
& +\sum_{n=0}^{\infty}\sum_{m=0}^{n} (2\omega^2\, \nu/\gamma)\, T_{nm} \\
& -\sum_{n=0}^{\infty}\sum_{m=0}^{n} \left[i \operatorname{tg}\alpha (n-m+1) Q_{n+1,m}(i \operatorname{cotg}\alpha) \right. \\
& \left. / Q_{nm}(i \operatorname{cotg}\alpha) \right] T_{nm} \quad ,
\end{aligned}
\tag{8.24}
$$

$$
T_{nm} = -k\, Q_{nm}(i \cot\alpha)(A_{nm} \cos m\lambda + B_{nm} \sin m\lambda) P_{nm}(\sin u) \quad ,
$$

where

g_A = the gravity anomaly,

ν = the ellipsoidal radius of curvature in the prime vertical,

γ = an average value of gravity,

k = the gravitational constant,

ω = the earth's rotation rate,

P_{nm} = the associated Legendre functions,

Q_{nm} = the Legendre's function of the second kind,

α = the angular eccentricity of the ellipsoid that passes through the point at which V is being computed,

u = the reduced latitude of the point at which V is being computed with respect to the ellipsoid passing through that point,

A_{nm}, B_{nm} = spheroidal potential coefficients referred to a specific ellipsoid,

i = $(-1)^{\frac{1}{2}}$.

It is noted that the effort needed to compute one gravity anomaly observation equation for the above model (i.e., to compute the partial derivatives and the constant term) would require many times the effort needed to compute one observation equation in conjunction with the corresponding spherical harmonic expansion. In fact, the complexity of the two models is so disparate that hardly any meaningful guess could be made in this respect. In addition, the initial values of the spheroidal harmonic coefficients would have to be computed beforehand from their spherical counterparts according to equation (24) of [Rapp, 1972']; this need arises from the fact that although the present equation (8.24) is linear in the parameters, the satellite altimetry model would not have this property and the best available values of the coefficients would be needed for the combination solution. Such a computation is quite complicated in itself. However, it would be performed only once. A similar statement can be made with regard to the conversion of the adjusted spheroidal harmonic coefficients back to the spherical ones as depicted in equation (25) of the same reference. In [Hotine, 1969], these two equations bear the numbers (22.59) and (22.60), respectively. Some of the closing remarks in [Rapp, 1972'] read:

> "Such a procedure would be more complicated than what is being done now. It would, however, reduce the convergent series problem to some extent. However, as less concern is attached to the convergence problem recently, it would appear questionable if the additional trouble in using spheroidal harmonics is worthwhile."

In discussing the convergence problems so far, we have been mainly concerned with the theoretical potential coefficients and their original physical meaning. However, at the end of the previous section, it has been pointed out that in a practical L.S. adjustment process, these coefficients have to a certain extent the role of regression parameters. This point will be raised again in one of the following concluding remarks.

We shall now briefly summarize a few findings made during this discussion, and attempt to draw a practical conclusion. We have observed that on the geoid itself -- which is of main concern -- the infinite series expressing the potential are formally divergent whether the spherical or the spheroidal harmonics are used. Thus the theoretical problems outlined at the beginning of both Sections 8.1 and 8.2 cannot be completely resolved. However, the theoretical inconvenience caused by the formal divergence in the series is believed to start with some high degree and order terms of the spherical harmonic expansion and with even higher degree and order terms of the spheroidal harmonic expansion of a given function.

It seems safe to infer that the level of truncation in the current applications at AFGL is such that the theoretical problems associated with the spherical harmonic expansion of the potential, gravity anomalies, etc., could hardly be noticed in practice.

The practical difficulties connected with an eventual adjustment process in terms of spheroidal harmonics could often reach prohibitive levels; in particular, the computer time requirements would most likely be many times higher than they are at the present time (the present software packages make use of the spherical harmonics).

Another practical consideration points to the fact that the potential coefficients in actual reductions serve, at least to a certain extent, in the role of regression parameters. That is to say that their task has become to make an adjustable function fit observations that have in general an unequal precision (and often an uneven distribution) and, subsequently, to make predictiors of that function -- or of a different function -- at arbitrary points, e.g. at grid points. These parameters would not have their original physical meaning even if the series were formally convergent and the observations happened to be errorless (a similar "regressionalization" would happen to parameters in another system because the two systems would be equivalent in the common region of convergence). On the other hand, since the mathematical model using the spherical harmonic coefficients is realistic and has many attractive attributes, and since the number of parameters is assumed to have a good relationship to the number and distribution of the observations, these parameters are suitable for many practical purposes and there is no urgent need to replace them with another set of regression parameters.

We may conclude that the use of spheroidal harmonics does not appear to be economically justified in connection with the actual data reductions at AFGL. The improvements in the practical outcome of such reductions would most likely be small. One of the chief factors responsible for it -- alone or in combination with other factors -- is the truncation of the initially infinite series. This conclusion is corroborated by Morrison [1969], who states in the closing paragraph:

> "Using ellipsoidal, or even highly exotic coordinate systems is not a solution to the problem. Truncated expressions in any other coordinate system would be equivalent to using spherical harmonics, and they would be far less convenient."

9. POINT MASS TECHNIQUE ANALYSIS

9.1 General Considerations

This analysis is based to a large extent on the report |Needham, 1970| which, for the purpose of the present chapter, will be abbreviated as |N|. In several instances, however, our approach will deviate from the one presented in |N| and it will be pointed out in the text. The main purpose of both approaches is of course identical; it is thus proper to make a few quotes from the chapter "Conclusions and Recommendations" of |N|, summarizing some fundamental features of this concept. On page 172 we read:

> "The point mass technique provides a simple method of using localized areas of detailed gravity information to add fine structure to a geopotential model that is based on spherical harmonic coefficients."

We may add that the detailed information considered in this analysis will consist not only of gravity data, but also of satellite altimetry data. In agreement with |N|, we shall introduce a model that will result in a true densification of the detail without distorting the long wavelength features in the original model formulated in terms of the spherical harmonics.

The fact that only long wavelength features are to be determined from the sphericalharmonic model stems from low degree and order harmonic coefficients (e.g. 14, 14 or 20, 20, etc.) envisioned in this model. The short wavelength features can be determined from locally introduced point masses as adjustable parameters. Such a combined approach can be used for the global representation

of the geoid as well as for describing its more detailed fluctuations. The main assets of this approach are its economy and its flexibility. In particular, in areas where an accurate geoid determination is not required or cannot be accomplished due to the lack of data, no point masses would be introduced; the geoid in such areas would accordingly be described by potential coefficients alone. This means that on the whole, the number of parameters could be kept relatively small (the complete set of parameters could consist of the above coefficients and magnitudes of point masses, one new parameter per point mass). The point masses would be employed in areas where, for various reasons, a more detailed knowledge of geoidal features may be required.

We may better appreciate advantages of the combined approach if we consider the task of describing certain short wavelength geoidal features over one limited area. In such a case, an adjustment solely in terms of potential coefficients would be wasteful since a great number of higher order and degree coefficients would have to be introduced just as if the knowledge of detailed features were required all over the globe. This presupposes that the observations provide a sufficient global coverage, in order to prevent the matrix of normal equations from being singular or severely ill-conditioned. The combined approach, on the other hand, would result in bringing into an adjustment only a relatively small number of point masses at appropriate locations. These newly added parameters would have the beneficial effect of determining the detailed fluctuations of the geoid surface over the selected area and nowhere else. The inclusion of an enormous number of parameters in an adjustment would thus be avoided. This is corroborated

by a statement on page 173 of |N|:

> "The point mass technique can be used to model fine detail in localized areas that would be completely impractical to represent by an ordinary spherical harmonic model. Such fine detail could be represented by anomaly data but computations based on point mass sets are far more simple and efficient than comparable computations based directly on gravity anomalies."

An apparent theoretical difficulty associated with the point mass concept is that there is an arbitrary number of point mass models generating a given potential. The comment made to this effect on page 173 of |N| reads as follows:

> "The point mass method has [a] disadvantage that a given set of gravity observations could be used to develop a theoretically unlimited number of different point mass models of the geopotential. There seems to be no difficulty in establishing a model that is accurate enough for practical applications, but it would be difficult to defend a specific model as the optimum for use in theoretical studies."

It is clear, then, that the solution to the problem has to be found by means of practical requirements and considerations.

The above problem has been treated in |N| under two stipulations, the first being that the gravity anomaly obtained in the point mass approach should agree as closely as possible with the gravity anomaly obtained via the surface density layer approach; a point mass and a density layer are considered at the same depth beneath the geoidal surface or, to a good approximation, beneath the surface of the reference ellipsoid. At the same time, the gravity anomaly at a given point should depend mostly on the point mass below it and the correlation with the point masses further away should

decrease very rapidly (second stipulation). A satisfactory solution to this problem as reported in [N] is the following. The region where the detailed features are sought is divided into equal area blocks of appropriate size, depending on the desired resolution. Each block, represented by a point mass at its center, is located at a certain depth underneath the surface of the reference ellipsoid. The depth which offers a good compromise with regard to both stipulations is about 0.8 of the block's side. It agrees reasonably well with the choice found e.g. in [Balmino, 1972], where this ratio is approximately 0.65. In this manner, the number and specific locations of the point masses are stipulated beforehand, while the magnitudes of the point masses are the parameters solved for in a least squares adjustment.

We now proceed to list the most important specific features of this analysis and their deviations from the approach presented in [N]; subsequently, they will be justified or explained in more detail. Full derivations will appear in some of the following sections. Perhaps the most important deviation from [N] is the notion that the basic surface described in terms of the spherical harmonic potential coefficients is not considered fixed. Instead, these coefficients may receive corrections in an adjustment of the observations that also serve to determine the magnitudes of point masses at given locations.

Another specific feature of this analysis is computational and practical in nature. While paying the price of introducing approximations that would be completely valid only under certain ideal conditions, we may significantly alleviate the burdensome computer storage requirements.

In fact, if a computer with extremely large storage capabilities were available allowing a simultaneous adjustment of many potential coefficients and point mass magritudes, the need for any approximation would be felt less urgently. In the absence of such a computer, however, we may have to adjust the coefficients and the point masses separately. Since the solution should ideally correspond to a simultaneous adjustment, this feature will be sometimes referred to as a "quasi-simultaneous" adjustment. In addition to the computer storage savings, it will also result in significant savings in terms of the computer run-time, especially if the number of parameters in both groups is large.

The treatment of constraints associated with the point masses represents another deviation of the present analysis from $|N|$. In this respect, we shall distinguish between the constraints applied differently, and the new constraints that do not appear in $|N|$ at all. The former group of constraints will be considered as "local" and the latter, as "global".

We shall now list one last deviation from $|N|$ which has been already mentioned at the beginning of this section. In particular, the scope of the point mass technique can be broadened by allowing for the adjustment of geoid undulations, derived from satellite altimetry observations. We recall that in the original model, only the gravity anomalies have been considered as observables.

We next proceed to a brief discussion of the important points in our analysis that have just been presented. First described will be the constraints because this feature will play a significant part in justifying

the adjustment of potential coefficients. The other features will be discussed according to the order in which they have been listed.

The constraints present in |N| assure that the point mass solution does not induce any change, with respect to the spherical harmonic model, in the total volume and the total mass considered in individual volume regions on the globe, subtended by certain equal area blocks. The side of these blocks has been reported such that it corresponds to the shortest half-wavelength of the spherical harmonic resolution. In this respect, an important deviation of our analysis from |N| resides in the proposition that the above volume regions should correspond to blocks whose side equals one whole shortest wavelength of the spherical harmonic resolution (it could also be more). Accordingly, if no constraints were added in our approach, the number of the volume regions and thus the number of these "local constraints" would amount to approximately 25% of their number in |N| . However, it is further suggested that in order to assure a smooth transition between individual blocks, similar constraints should be introduced also in overlapping volume regions. The number 25% will thus be ultimately replaced by 50%. Since both the "local volume" and the "local mass" are constrained to be unaffected by the inclusion of the point masses, the "global volume" and "global mass" of the spherical harmonic model will be automatically preserved (this statement is exact with respect to the mass, and almost exact with respect to the volume).

We shall now discuss the possibility of using some new constraints that do not appear in |N|. In order to preserve the coincidence of the coordinate origin and the global center of mass, we could introduce three constraints (their three counterparts in the spherical harmonic model remove the first degree harmonics by holding the three pertinent parameters equal to zero, namely $C_{10} = C_{11} = S_{11} = 0$). The point mass constraints preserving this coincidence were used by Balmino |1972|; however, no local constraints were present in his special type of adjustment. We could further apply two new constraints that would keep the earth's axis of rotation coinciding with the Cartesian z - axis (in the spherical harmonic model this is achieved by enforcing $C_{21} = S_{21} = 0$, i.e., by applying two additional constraints). The point mass constraints of this kind have appeared neither in the above reference nor in |N|. Finally, we can consider one last new constraint that would preserve the orientation of the Cartesian x, y axes with respect to the earth's principal axes of inertia (in the spherical harmonic model, this is achieved by enforcing a given value of S_{22}, i.e., by applying one additional constraint). We have thus presented six "global constraints" that are sufficient to preserve the original coordinate system after the inclusion of the point masses. It is of course a well-known fact that six constraints are needed to define the coordinate system (three to define its position and three to define its orientation).

The importance of preserving the coordinate system as defined in the spherical harmonic model becomes more evident in connection with the short arc mode of satellite altimetry. The satellite-related quantities in this model depend to a large extent on the total mass, on some low degree

and order potential coefficients, and on the coordinate system associated with the given state vector components (in this analysis it has been the earth fixed system). We have seen earlier that the total mass is completely unaffected by the point masses. The same can be said, at least from the practical standpoint, about the low degree and order potential coefficients (they identify the long wavelength features of the gravity field). If the coordinate system is also preserved, the adjusted satellite parameters will remain consistent with the new gravity field representation, containing also the point masses. In other words, the satellite parameters need not be considered at all in conjunction with the point masses. This is of course a desirable property that helps the adjustment model to remain relatively simple. We shall next discuss whether or not, in the present context, the above six constraints are indeed necessary in order to prevent the coordinate system from changing due to the point masses.

The first proposition that must be confronted in this situation indicates that coordinate shifts and rotations are in fact very long wavelength features. As we have seen earlier, they are associated with the very low degree and order coefficients C_{10}, C_{11}, S_{11}, C_{21}, S_{21}, and S_{22}. Such features should not be influenced by the presence of the point masses if the local constraints are correctly applied. We may indeed suspect that the local mass constraints alone are sufficient to preserve the coordinate system unchanged. This can be confirmed as follows. We have seen that the equal area blocks associated with the local constraints have sides equal to the shortest wavelength of the spherical harmonic resolution. As the wavelengths increase, a number of point masses introduced beneath one such block

behave more and more like a simple point mass considered beneath the center of the block. However, due to the local mass constraints, these new (or aggregated) point masses act in fact as if they all had zero magnitude. It is to be re-emphasized that this property holds better for longer wavelengths. For instance, the point masses are expected to excercise very little influence upon the wavelengths that are four -- or more -- times greater than the above shortest wavelength. This proposition is reinforced by the consideration of additional constraints in overlapping regions. It is then clear that the coordinate system would be the least affected by the point masses (besides the total mass which is of course not affected at all). We may thus conclude that for practical purposes, the coordinate system would be preserved without the application of the six constraints described earlier. However, no harm can be done if they are introduced for academic or other reasons. In any event, their number is small and their form is simple. As in the case of other constraints, they could be implemented as observation equations and assigned very large weights. The most important outcome of the discussion so far has been the recognition that the coordinate system can be preserved without special constraints due to the point masses, and that the only indispensable constraints in the adjustment of the point masses are of the local type. They are about one half in number when compared to those presented in $|N|$.

We now turn to the reasoning behind the proposed treatment of the spherical harmonic potential coefficients. The practical advantages of using adjustable potential coefficients as opposed to holding them fixed at some appropriate nominal values are the following. First, one may be interested

in computing the coefficients that best fit the available data, be it for the sake of comparisons or for other reasons; and second, the areas of higher data density are allowed to play their role in determining the "smoothed geoid" in those areas, which means that the detail added via point masses will have random -- rather than systematic -- character. In other words, the magnitudes of point masses will be relatively small and their sign will vary fairly randomly. Similar conclusions will hold also with respect to the short wavelength features of geoid undulations. But this is perfectly compatible with the constraints mentioned earlier with regard to the locally unchanged mass and the locally unchanged volume. It is thus plausible that these constraints will introduce only insignificant distortions in the actual residuals when compared to the residuals that would be obtained from an unconstrained adjustment. If, on the other hand, a given set of fixed potential coefficients incompatible with the "observed geoid" in areas of higher data density were used, then, due to these constraints, the point masses would not be able to accommodate the resulting systematic differences and the computed geoid would not fit the observations. Some systematic distortions in the residuals as well as in the relatively long wavelength features would remain and they could even influence some of the short wavelength detail. Necessarily, the a-posteriori RMS error would be larger than in the suggested approach.

The proposed "quasi-simultaneous" adjustment can be conceived along the following lines. First, an adjustment of a given set of potential coefficients could be performed using all the available data (e.g. mean gravity anomalies, satellite altimetry measurements). This would yield a surface

that should ideally coincide with a smoothed geoid. In areas of higher data density, the residuals produced by this adjustment could then represent new observations in a separate adjustment, in which the magnitudes of point masses would be determined. In this adjustment, the wavelengths of geoid features should be sufficiently short, adding detail to the "smooth geoid" obtained previously. The final results would closely approximate a simultaneous adjustment of both the potential coefficients and the point masses, task that would otherwise entail enormous core space requirements. This approximation is better when the observations are independent and have the same precision, when they are evenly distributed, when they are sufficiently dense on the global scale, etc. The point mass constraints discussed previously would of course enter the second adjustment.

Finally, we observe that the form of the gravity anomaly observation equations in a simultaneous adjustment would be the same as the form of the observation equations derived for the spherical harmonic potential coefficients alone, except that terms resulting from the products between a new set of parameters (point mass magnitudes) and a corresponding set of partial derivatives would join the existing terms. The same can be said with regard to the observation equations of satellite altimetry. We have seen that in the "quasi-simultaneous" adjustment, the two parts of the observation equations are separated. In either case, the partial derivatives appearing in the formulas for variance-covariance propagation have the same form as the corresponding expressions in the observation equations, be it for gravity anomalies or satellite altimetry. However, caution should be exercised when expressing final variances in the suggested "quasi-simultaneous" approach. Clearly, the variance at a given point obtained through

the variance-covariance propagation in the second adjustment (i.e., in terms of the point masses) would be too optimistic. It must be augmented by the variance attributable to the potential coefficients, computed in the first adjustment. More about the "quasi-simultaneous" adjustment and the variance-covariance propagation will be said in Section 9.4.

Other examples related to the point mass concept may be found in the geodetic literature. For example, Balmino |1972| computed a set of point masses through a least squares fit to the SAO field. The observation equations were formed in terms of gravity anomalies generated essentially by point masses alone, and were required to fit the gravity anomalies derived from the SAO potential coefficients. In other words, the point masses were adjusted to fit a given set of potential coefficients -- rather than observations. The method of |Balmino, 1972| is clearly different from the approach presently described, whose prominent feature is the fitting of both spherical harmonic coefficients and point mass magnitudes to observed or mean gravity anomalies and to geoid undulations derived from satellite altimetry observations.

9.2 Spherical Harmonic Resolution

The number of spherical harmonic potential coefficients in a model developed for degree and order (k, k) is $(k + 1)^2$. Of these coefficients, five are usually held fixed at the zero value and a sixth may be constrained to a given value (these six paramaters define a coordinate system). For the purpose of approximately determining the minimum number of observations needed to compute the unknown potential coefficients, it will not be necessary to distinguish between $(k + 1)^2$, $(k + 1)^2 - 5$, or $(k + 1)^2 - 6$; the practical outcome of such considerations would be about the same in all three cases, especially for larger values of k. Accordingly, the number of parameters in the spherical harmonic model is adopted in this section to be $(k + 1)^2$.

We shall now regard the earth as a unit sphere divided into discrete blocks of equal area ($\tilde{\Delta S}$). Each block represents an observation (e.g. a gravity anomaly referred to the center of the block). The obvious question arises: What should be the central angle ($\tilde{\Delta\theta}$) identifying the side of these blocks that would just allow for the determination of all the parameters? Clearly, the number of blocks must be $(k + 1)^2$, hence

$$\tilde{\Delta S} \equiv \tilde{\Delta\theta} \times \tilde{\Delta\theta} = 4\pi / (k + 1)^2 \quad . \tag{9.1a}$$

We thus have

$$\tilde{\Delta\theta} = 2 (\pi)^{\frac{1}{2}} / (k + 1) \quad , \tag{9.1b}$$

or

$$\tilde{\Delta\theta}^0 = 203.1^0 / (k + 1) \quad . \tag{9.1c}$$

The formula (9.1c) appeared on page 74 of |N|, but the reasoning behind it was different from ours. In particular, the formula was justified by the relation appearing on the top of that page, which, however, led to a different number of local constraints as compared to the number we shall derive.

In practice, the number of observations is usually larger than is necessary to just determine all the parameters. Accordingly, the block sides would be in general smaller than indicated by equations (9.1), if the approach with equal area blocks were adopted. If the equal area blocks were to be replaced by geographic blocks of the same nominal dimensions (e.g. $5^0 \times 5^0$, etc.), the number of the new blocks would increase significantly; in particular, near the poles, one equal area block would be equivalent in size to several geographic blocks. This is in fact the usual format in which the gravity anomaly data is given (together with standard errors as a precision indicator for each geographic block). The number of observations is thus large, but their identification and treatment are simplified so that this procedure is quite suitable for practical computations. Since such practical considerations do not affect the present analysis, they will henceforth be disregarded.

We shall now confine our attention to determining the shortest half wavelength of the spherical harmonic resolution. Needed for this purpose are the formulas giving T (the disturbing potential), N (the geoid undulation), and Δg (the gravity anomaly) in terms of the conventional spherical harmonics. They are written below in the spherical approximation, complete through the degree and order (k, k):

$$T \simeq G R \sum_{n=2}^{k} \Delta S(n) \quad , \tag{9.2a}$$

$$N \simeq R \sum_{n=2}^{k} \Delta S(n) \quad , \tag{9.2b}$$

$$\Delta g \simeq G \sum_{n=2}^{k} (n - 1) \Delta S(n) \quad , \tag{9.3}$$

where

$$\Delta S(n) = \sum_{m=0}^{n} (\Delta C_{nm} \cos m\lambda + \Delta S_{nm} \sin m\lambda) P_{nm} (\sin \bar{\phi}) \quad , \tag{9.4}$$

and where G, R are the mean values of gravity and of the earth's radius, respectively. Equations (9.2b), (9.3), and (9.4) appeared essentially in the same form in |Rapp, 1971|, page 3 (both the coefficients ΔC_{nm}, ΔS_{nm} and the associated Legendre functions P_{nm} were normalized). Equations (9.2a) and (9.2b) are equivalent, in the sense that

$$N \simeq T / G \quad ,$$

which follows from Bruns formula.

The surface spherical harmonics $\Delta S(n)$ are called, in general, tesseral harmonics. They divide a unit sphere into compartments in which they are alternately positive and negative. For given n and m, $\Delta S(n)$ has n - m zeros along a meridian, and 2m zeros along the equator or any parallel (for the moment, we shall consider only the equator). The maximum number of zeros in the interval $-\frac{1}{2}\pi \leq \bar{\phi} \leq \frac{1}{2}\pi$ is k, and it occurs when n = k and m = 0. The zeros are spaced fairly regularly. They give rise, within 180^0 in latitude, to approximately k half-waves associated with ΔC_{k0} and representing the resolution capability. Accordingly, the shortest half wavelength along a meridian is

$$\kappa_1^0 \simeq 180^0 / k \quad .$$

On the equator, the maximum number of zeros in the interval $0 \leq \lambda < 2\pi$ is $2k$, and it occurs when $n = m = k$. The zeros are spaced regularly and the waves are associated with ΔC_{kk}, ΔS_{kk}. We now have 2k half-waves within 360^0, or k half-waves within 180^0, similar to the previous case. The shortest half wavelength along the equator thus is

$$\kappa_2^0 = 180^0 / k \quad .$$

Considering the earth as a sphere of radius R, the shortest wavelength in linear units along a meridian or along the equator is approximately $2\kappa_2 R$. Much of the above discussion has been based on pages 25 and 26 of |Heiskanen and Moritz, 1967|, abbreviated as |HM|.

It may appear that as one proceeds away from the equator, the resolution in longitude improves since for a constant κ_2, the corresponding linear units decrease. However, this added capability is ficticious because the seemingly finer and finer resolution is severely dampened by the factor P_{kk} (sin $\bar{\phi}$) as $\bar{\phi}$ approaches the poles. In order to see this we merely have to consider

$$P_{kk} (\sin \bar{\phi}) = \left| 1 \times 3 \times \ldots \times (2k - 1) \right| \cos^k \bar{\phi} \quad ,$$

which decreases rapidly in the situation just described, due to the k th power in cos $\bar{\phi}$. This outcome is plausible since the resolution in one area cannot be arbitrarily improved simply by changing the orientation of the coordinate system, everything else remaining the same.

The well-known property that $\Delta S(n)$ -- or its equivalent -- is invariant with respect to coordinate rotations stems e.g. from the comparison of equations (1 - 66) and (1 - 71) in |HM|. Using our present notations, we replace these two formulas respectively by

$$N / R \simeq \sum_{n=2}^{k} \Delta S(n) \quad ,$$

$$\Delta S(n) \simeq \left|(2n + 1)/(4\pi)\right| \iint_{\sigma} (N/R) P_n(\cos\Psi) d\sigma \quad , \qquad (9.4')$$

where σ represents the surface of a unit sphere, and $d\sigma$ is its element. The angle Ψ denotes the spherical distance between the point where $\Delta S(n)$ is sought (it has the spherical coordinates $\overline{\phi}$, λ) and the point associated with $d\sigma$. The value given by (9.4') is clearly invariant with respect to the orientation of the coordinate system, i.e., with respect to coordinate rotations that might have been otherwise suspected to affect the outcome of (9.4). We thus conclude that $\Delta S(n)$ and the resolution that it entails are indeed invariant with respect to such rotations. As a by-product, we see immediately that degree variances also have the same property; in this particular case, the degree variance is the average square (over the unit sphere) of the harmonic $\Delta S(n)$ of degree n.

We have examined the resolution possibilities along a meridian and along the equator. In view of the last paragraph, we may state that they are the same everywhere; there is no preferred location or direction. Denoting the shortest half wavelength of the resolution by $\Delta\theta$ (previously it was denoted

as κ_1 or κ_2 according to the direction considered), we conclude that

$$\Delta\theta^0 \simeq 180^0 / k \quad .$$

This relation has been already encountered in Section 8.1. It was used by Rapp [1974], page 39, in asserting that a 15^0 discrete anomaly block field is roughly equivalent to a spherical harmonic expansion to degree 12. This indicates that the equal area blocks he was using could have been constructed in agreement with

$$\Delta S \equiv \Delta\theta \times \Delta\theta \quad , \tag{9.5a}$$

where $\Delta\theta$ is again a central angle, given now as

$$\Delta\theta \simeq \pi / k \quad , \tag{9.5b}$$

or

$$\Delta\theta^0 \simeq 180^0 / k \quad . \tag{9.5c}$$

An approximate, but very simple and efficient procedure for computing such blocks in practice was used by Rapp [1974] in terms of geographic coordinates. It is symbolized by the following two equations:

$$\Delta\phi \simeq \Delta\theta \quad , \tag{9.6a}$$

$$\Delta\lambda \simeq \Delta\theta / \cos\phi \quad , \tag{9.6b}$$

where $\Delta\phi$, $\Delta\lambda$ may often be rounded to the nearest degree.

We have now two methods at our disposal for computing the equal area blocks that would yield the values of spherical harmonic coefficients without an adjustment (or with almost no adjustment). The methods are represented respectively by equations (9.1), described in [N], and by equations (9.5), used by Rapp [1974] in the practical form (9.6). The two methods are not contradictory since the block sizes differ by only small amounts. To demonstrate this, we list, in Table 9.1, the sides $\Delta\tilde{\theta}^0 = 203.1^0/(k+1)$ of (9.1c) and $\Delta\theta^0 = 180^0/k$ of (9.5c) in conjunction with a few values of k, meaningful in practical applications (the value k = 14 appears to be a suitable candidate for real data reductions with point masses). The table shows that from k = 8 on, the sides respresented by (9.5c) are slightly smaller; this indicates a very small increase in the number of blocks and, therefore, of observations. The differences are more of theoretical than of practical interest. For instance, the value k = 14 demonstrates only about 5% difference in sides.

k	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$\Delta\tilde{\theta}^0$	29.0	25.4	22.6	20.3	18.5	16.9	15.6	14.5	13.5	12.7	11.9	11.3	10.7	10.2	9.7
$\Delta\theta^0$	30.0	25.7	22.5	20.0	18.0	16.4	15.0	13.8	12.9	12.0	11.3	10.6	10.0	9.5	9.0

TABLE 9.1

Sides $\Delta\tilde{\theta}$ and $\Delta\theta$ for a few useful values of k

Since equations (9.5) result in a slightly more conservative approach, they shall be used throughout the pertinent analysis. The advantages of such a choice are its simplicity and intuitive appeal, as well as the angular equivalence between the block sides and the shortest half wavelength of the spherical harmonic resolution. In |Rapp, 1972|, page 16, this shortest half wavelength is called "the angular half wavelength of the information", but it is related to the block sides in the same way as in our approach (see page 5 of this reference). The number of blocks is given as

$$b = 4\pi / \Delta S \equiv 4\pi / (\Delta\theta \times \Delta\theta) \quad ; \qquad (9.7a)$$

the number of blocks constructed according to the practical equations (9.6) may be slightly different from b. If p identifies the number of parameters (here the potential coefficients), we may write

$$b \simeq (k + 1)^2 \simeq p \quad . \qquad (9.7b)$$

The quantity b in (9.7b) may be slightly larger than $(k + 1)^2$ which, in turn, may be slightly larger than p if this number excludes the fixed parameters.

Finally, a thought should be given to the relationship between the observational density and the shortest half wavelength. The notion of the shortest half wavelength may now be considered both in the context of a spherical harmonic adjustment alone, and in the context of the spherical harmonic model with point masses; as we have seen, the latter model could be treated in a theoretical simultaneous adjustment or in a practical quasi-simultaneous adjustment. In a spherical harmonic model, the shortest half wavelength was given in (9.5b) or (9.5c). When the point masses are

superimposed on this model, the resolution possibilities will necessarily improve. In fact, in analogy to the gravity anomaly blocks, we shall consider equal area blocks associated each with a point mass. The shortest half wavelength will be then represented by the blocks' sides. In either case, however, one observation per block associated with the shortest half wavelength would yield the parameters with little or no filtering. Any error in the observations would affect the shortest waves directly, to the fullest extent.

In analogy to a one-dimensional case, it is reasonable to at least double the minimum number of observations in each dimension. Thus, in general, each shortest half-wave would be determined by at least two observations and the whole wave, by at least four observations. The number of observations on the sphere (i.e., in the present two-dimensional case) would thus be at least four times as large as the number of equal area blocks corresponding to the shortest half wavelength. Conversely, based on the observational density, we can determine the degree and order beyond which an adjustment in spherical harmonics should not be carried out, or the number of point masses in a given area which should not be exceeded. For instance, a $5^0 \times 5^0$ anomaly block field could yield a good determination of potential coefficients up to and including the degree and order (18, 18), the corresponding shortest half wavelength being 10^0. The anomaly block field is assumed to consist of equal area blocks on the global scale, containing each a (mean) gravity anomaly . In the case of global coverage by geographic blocks, a similar conclusion would apply (the "extra" observations would be present mostly in the polar regions, at least nominally, but not near the equator).

We may consider a $1^0 \times 1^0$ anomaly block field in a similar fashion. In this case, a good determination with regard to point mass magnitudes could be made in $2^0 \times 2^0$ equal area blocks, or in larger blocks, each associated with a point mass. These $2^0 \times 2^0$ blocks would of course extend only over the region of interest. The point masses would be superimposed on some spherical harmonic model, such as the one represented by the degree and order (14, 14). Without the point masses, the spherical harmonic model that would yield a comparable resolution in the region of interest -- and everywhere else -- would be represented by the degree and order (90, 90). This would entail enormous computational difficulties, since some 8281 parameters would have to be resolved. However, a $1^0 \times 1^0$ global anomaly field is nonexistent; gravity anomalies of such a density are accesible only in selected regions of the earth. It is then necessary to have recourse to the point masses as discussed above, or to an equivalent representation, in order to achieve the desired local resolution.

Similar considerations could be also made with respect to geoid undulations derived from satellite altimetry observations. Since the altimetry measurements are not in general located in an equal area grid (or any other regular grid), the data coverage would have to be assessed very carefully if one wants to retrieve optimum results from an adjustment. However, an important point to be made is that if the geoid undulations were indeed given in equal area blocks, the shortest wavelength characteristics, etc., would be similar to those encountered when considering mean gravity anomalies. In fact, from (9.2b), (9.3), and (9.4), we observe that the respective shortest wavelengths are the same, but due to the factor $(n - 1)$, Δg is a markedly less smooth function than N.

9.3 Some Properties of the Point Mass Model

We have seen that the depth/side ratio of point masses associated with equal area blocks was determined in |N| under two stipulations, both related to gravity anomalies. The formula giving the gravity anomaly in terms of the point masses was presented in |N| as equation (2.19), while (2.21) identified a similar formula corresponding to the density layer approach. Under the first stipulation, the depth/side ratio (denoted here as d/s) should be such that the results computed by the above equations (2.19) and (2.21) are nearly identical.

To examine when and how well this task is accomplished, the kernel of either equation was computed for individual blocks, starting with the block directly under the computation point and proceeding outwards. If the two kernels showed a good agreement for all the blocks of any significance, the gravity anomalies computed by the two methods would then also agree. The results of the comparisons in blocks were presented in Table 1 of the same reference. They appear good for d/s = 0.8 and keep improving for higher and higher ratios. However, starting already with the third block from the computation point, the results were seen to be in excellent agreement for all the listed ratios. With regard to the first stipulation, it may be concluded that the point mass representation is sufficiently close to the density layer representation if d/s is at least 0.8.

The second stipulation implies that the gravity anomaly should depend almost entirely on the point mass directly below it. The choice of d/s that would satisfy this requirement is greatly facilitated upon using

the point mass kernel mentioned in the preceding paragraph. This kernel illustrates the dependence of the gravity anomaly at the computation point upon individual point masses associated with blocks at various locations. The pertinent results appear in Table 1 and Table 3 of |N|. The quantities of interest from these two tables are rearranged and reproduced here as Table 9.2, where ℓ denotes the lateral distance between the computation point and the center of a point mass block. A satisfactory outcome is obtained with d/s = 1.0, but significant improvements take place with decreasing d/s. A visual display of the results from Table 9.2 is presented in Figure 9.1.

ℓ	0.00 s	0.25 s	0.50 s	0.75 s	1.00 s	2 s	3 s	4 s
d/s = 0.8	1.523	1.321	0.920	0.578	0.356	0.065	0.017	0.004
d/s = 1.0	0.969	-----	-----	-----	0.331	0.075	0.022	0.007
d/s = 2.0	0.234	-----	-----	-----	0.165	0.077	0.034	0.015

TABLE 9.2

The kernel of the point mass model for various lateral displacements ℓ and for different ratios d/s

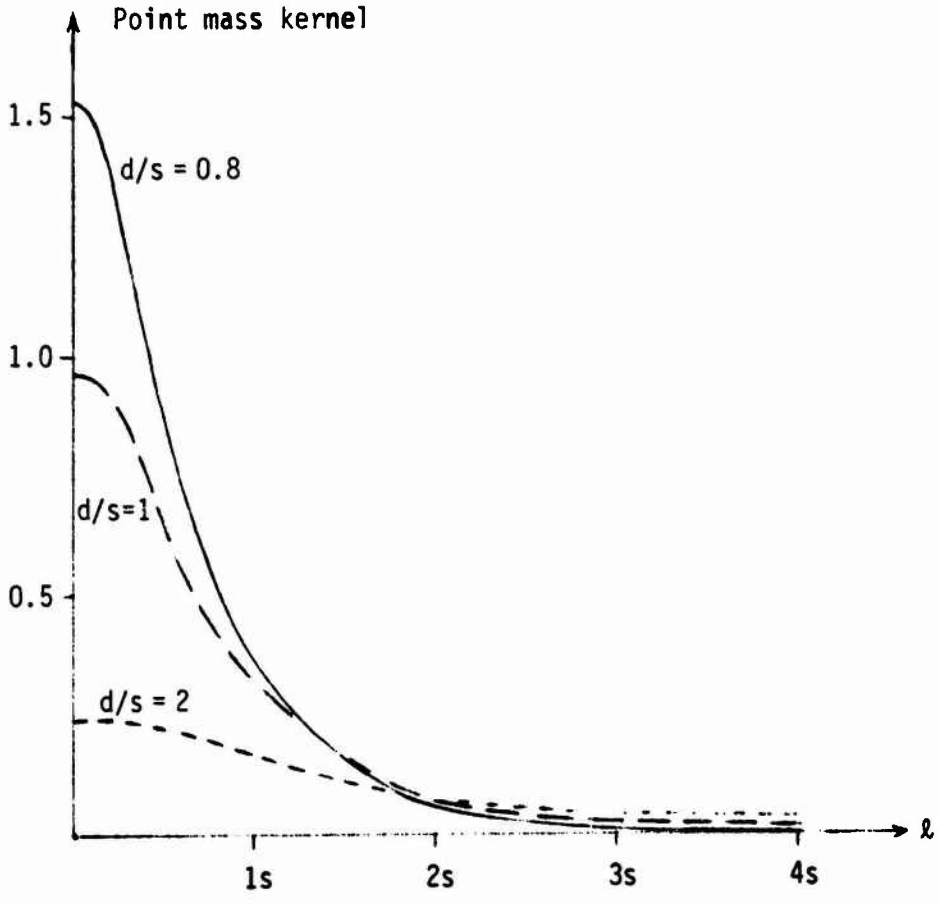


FIGURE 9.1

Visual representation of the results from Table 9.2

Let us now consider an example offered by Table 9.2. We can deduce that if the point mass kernel is adopted as a yardstick and if d/s = 0.8 is accepted, the dependence of the gravity anomaly on the single point mass block directly under the computation point accounts for approximately one half of the dependence on all the other blocks of the globe taken together. The point masses in all the possible blocks at lateral distances of three or more blocks from the computational point contribute less then 10% of the total

effect. Such a contribution would be completely negligible if the point masses at a lateral distance of at least four blocks were considered instead. The second stipulation is thus fulfilled very satisfactorily if $d/s \leq 0.8$. However, the first stipulation had led to the relation $d/s \geq 0.8$. The obvious conclusion, then, is that a suitable value for the point mass analysis is

$$d / s = 0.8 \quad . \tag{9.8}$$

The rapid decrease in correlation between the gravity anomaly and the point masses with the increase in lateral distances indicates that the point masses can be used to represent about the same information as the gravity anomalies. According to Section 2.1 of $|N|$, the point masses are interpreted as "anomalous masses" generating the distrubing potential of the normal gravity field. The corresponding gravity anomalies refer to the same field. The introduction of a relatively low degree and order spherical harmonic model will not change appreciably the above near equivalence between the gravity anomalies and the point masses. This could be explained by the fact that several anomaly blocks would be usually present within one shortest half wavelength of the spherical harmonic representation. The spherical harmonic resolution thus could not significantly affect the behavior of two neighboring anomalies. But since the point masses depend on just a few anomalies closest to them, and vice versa, their correlation would change very little. In fact, it is believed that the nearly one-to-one correspondence between the gravity anomalies and the point masses would actually improve under such conditions.

The immediate consequence of this correspondence is that as far as the shortest half wavelengths, etc., are concerned, the equal area point mass blocks have the characteristics similar to the equal area gravity anomaly blocks. In particular, the shortest half wavelength can be respresented by the side of the point mass blocks. Another consequence is little or no correlation between distant point mass magnitudes; since distant gravity anomalies are practically uncorrelated, the same should hold true with respect to the point masses. This is confirmed by the text and by an example on pages 86 and 87 of $|N|$, from which it follows that the correlation between the point masses located three or more blocks apart is indeed negligible.

From the foregoing discussion it appears that if we were interested, for some reason, in adjusting an predicting exclusively the gravity anomalies through the point mass model, no higher order surface would be necessary; such a surface, replacing the reference ellipsoid, is represented by a low degree and order spherical harmonic model. However, the situation changes drastically if the geoid undulations (N) are brought into the picture, whether as a part of the adjustment process or as the quantities to be predicted. In order to compute one geoid undulation, gravity anomalies are required everywhere on the globe. This is evidenced by the Stokes' formula where the integration of gravity anomalies times Stokes' function is extended over the whole unit sphere. If only a limited area within a cap of central angle Ψ_o around the computation point were included (it will be called a truncation cap Ψ_o), the accuracy in determining N could be severely impaired.

The same defect occurs if the geoid undulations are to be expressed in terms of the point masses, due to the close correspondence between the gravity anomalies and the point masses. To compute N, point masses all over the globe would have to be taken into account since a truncation cap Ψ_0 would yield about the same errors as would be expected if the gravity anomalies were considered instead. This task would be computationally very burdensome, if it could indeed be performed at all. We have seen that in practice, the point mass blocks are introduced only in a limited area. Under the usual circumstances, then, an attempt to consider the point masses as a means of expressing geoid undulations directly, referred to an ellipsoidal surface, could be completely invalidated. For instance, an average error in N in conjunction with Ψ_0 as large as 40^0 would still reach almost 10 m (see Figure 4 in $|N|$).

Fortunately, the above problem is substantially alleviated by the introduction of a higher order surface. Upon using the spherical harmonic model through the degree and order (14, 14), the truncation cap as small as 13^0 yields an average error in N of only about 0.6 m, as can be gathered from Figure 6 of the same reference. In order to obtain good values of adjusted or predicted undulations everywhere in the region of interest (i.e., including its periphery), the point mass blocks should ideally cover a proportionately larger area. We can further observe that an integration cap of $\Psi_0 \simeq 15^0$ leads to errors smaller than 0.5 m. If the average error in N is required smaller than 0.1 m, the integration cap is symbolized by

$$\Psi_0 \simeq 40^0 \quad . \tag{9.9}$$

This represents about a hundred-fold improvement when compared with the direct approach where no higher order surface is present. If the higher order surface were represented by the spherical harmonic model (k, k), where $k > 14$, such improvements would be even more significant. We can thus conclude that the use of a higher order surface minimizes the importance of the outer anomaly field, or, equivalently, of the outer blocks associated with point masses. As stated in [N], page 71, this is one of the key factors making the concept of a point mass representation computationally feasible.

As a practical remark, we may add that in spite of a higher order surface, a large number or all of the point masses present in the model have to be considered with every geoid undulation. This is especially true if the relation (9.9) should apply. By contrast, the point masses needed in the adjustment or predictions of gravity anomalies are relatively few. In particular, they comprise only the point masses that are within one or two blocks from the computation point. Such a cap is certainly much smaller than Ψ_0 associated with N. However, the computational disparity with regard to Δg and N is not as great as it may seem since in terms of the point masses, the expression for N will be shown to be much simpler than the expression for Δg.

9.4 Least Squares Adjustment without Constraints

9.4.1 General Considerations of Two Independent Sets of Parameters

The development in this part will serve later, in the adjustment of spherical harmonic coefficients and point mass magnitudes. The parameters are therefore considered in two groups which ideally should be mutually independent. We shall present two equivalent methods, of which the second will be applicable in practice; the first method will serve essentially as a guideline and for verifications. The notations bear resemblence to those used in [Blaha, 1975], pages 5 and 6, and they will be explained first.

We start with the adjustment model

$$L^a = F(X_1^a, X_2^a) \quad , \tag{9.10}$$

where the superscript "a" denotes adjusted quantities; the vector L^a contains adjusted observations and the vectors X_1^a, X_2^a represent the two groups of adjusted parameters. The relationship between the observables and the parameters is symbolized by F. We also have

$$L^a = L^b + V \quad , \tag{9.11}$$

where L^b is the vector of observations and V is the vector of residuals. The initial values of parameters in the first group are arranged in the vector X_1^0, so that

$$X_1^a = X_1^0 + X_1 \quad , \tag{9.12}$$

where X_1 contains the corrections to the contants in X_1^0. On the other hand, the initial values in the second group of parameters are taken as zeros; in

particular,

$$X_2^0 \equiv 0 \quad , \tag{9.13}$$

hence

$$X_2^a \equiv X_2 \quad . \tag{9.13'}$$

This indicates that the relationship between the observables and the parameters in the second group is considered linear. If the complete model (9.10) is linearized, it may be written in the following form:

$$L^b + V = A_1 X_1 + A_2 X_2 + L^0 \quad , \tag{9.14}$$

where the matrices A_1, A_2 are expressed as

$$A_1 = (\partial F / \partial X_1)_0 \quad , \quad A_2 = (\partial F / \partial X_2)_0 \quad ,$$

and the constant vector L^0 is formed according to

$$L^0 = F(X_1^0, X_2^0) = F(X_1^0, 0) \quad .$$

The weight matrix, associated with the observations, is defined as

$$P = \sum_{L^b}^{-1} \quad . \tag{9.15}$$

In the first method, the parameters in both groups are solved for in one simultaneous least squares adjustment. From (9.14), the observation equations are written as

$$V = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + L \quad , \tag{9.16}$$

where

$$L = L^0 - L^b \quad . \tag{9.16'}$$

The least squares condition $V^T P V$ = minimum leads to the following solution of normal equations, where the full rank is henceforth assumed whenever necessary for the regular matrix inversions:

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = -\begin{bmatrix} A_1^T P A_1 & A_1^T P A_2 \\ A_2^T P A_1 & A_2^T P A_2 \end{bmatrix}^{-1} \begin{bmatrix} A_1^T P L \\ A_2^T P L \end{bmatrix} .$$

We now consider the two parameter groups as completely independent, in the sense that

$$A_1^T P A_2 = 0 \quad .$$

The above solution thus becomes

$$X_1 = -(A_1^T P A_1)^{-1} A_1^T P L \quad , \tag{9.17a}$$

$$X_2 = -(A_2^T P A_2)^{-1} A_2^T P L \quad . \tag{9.17b}$$

The variance-covariance matrix for the parameters is composed of two diagonal blocks as follows :

$$\sum_{X_1, X_2} = \text{diag}\left(\sum_{X_1} , \sum_{X_2}\right) \quad , \tag{9.18a}$$

where

$$\sum_{X_1} = (A_1^T P A_1)^{-1} \, , \tag{9.18b}$$

$$\sum_{X_2} = (A_2^T P A_2)^{-1} \, . \tag{9.18c}$$

The adjusted observations and the predictions (at locations other than observations points) are given by the formulas that are very similar in form, namely

$$L^a = L^0 + \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} , \qquad (9.19)$$

$$L'^a = L'^0 + \begin{bmatrix} A'_1 & A'_2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} , \qquad (9.20)$$

where (9.19) is in fact (9.14) and (9.20) is the linearized form of

$$L'^a = F' (X_1^a , X_2^a) \quad .$$

Upon propagating the variances-covariances depicted in equations (9.18), from (9.19) we obtain

$$\sum_{L^a} = A_1 (A_1^T P A_1)^{-1} A_1^T + A_2 (A_2^T P A_2)^{-1} A_2^T \quad , \qquad (9.21)$$

while (9.20) yields

$$\sum_{L'^a} = A'_1 (A_1^T P A_1)^{-1} A_1'^T + A'_2 (A_2^T P A_2)^{-1} A_2'^T \quad . \qquad (9.22)$$

In the second method, a first adjustment is carried out ignoring the second group of parameters; meanwhile, the observations and the corresponding weights are the same as before. Denoting the new residuals as $\tilde{V}$,

we have

$$L^b + \tilde{V} = A_1 X_1 + L^0 \quad , \tag{9.23}$$

which corresponds to the adjustment model

$$\tilde{L}^a = F(X_1) \quad ,$$

where

$$\tilde{L}^a = L^b + \tilde{V} \quad . \tag{9.24}$$

The constant vector L^0 is unchanged with respect to the first method since

$$L^0 = F(X_1^0) \equiv F(X_1^0, 0) \quad .$$

From (9.23) we thus form the observation equations in the first adjustment as follows:

$$\tilde{V} = A_1 X_1 + L \quad , \tag{9.25}$$

where L is also the same as before; it is given in equation (9.16'). The least squares condition $\tilde{V}^T P \tilde{V}$ = minimum now results in the solution

$$X_1 = -(A_1^T P A_1)^{-1} A_1^T P L \quad ,$$

which agrees with (9.17a). The variance-covariance matrix for X_1 agrees with (9.18b); it is given as

$$\sum_{X_1} = (A_1^T P A_1)^{-1} \quad .$$

We also have

$$\tilde{L}^a = L^0 + A_1 X_1 \quad , \tag{9.26}$$

$$\tilde{L}'^a = L'^0 + A_1' X_1 \quad , \tag{9.27}$$

and

$$\sum_{\tilde{L}^a} = A_1 (A_1^T P A_1)^{-1} A^T \quad , \tag{9.28}$$

$$\sum_{\tilde{L}'^a} = A_1' (A_1^T P A_1)^{-1} A_1'^T \quad . \tag{9.29}$$

If we want to adjust an additional, independent group of parameters X_2 such that (9.13) applies, the residuals $\tilde{V}$ from the first adjustment will play the role of new, ficticious observations having the opposite sign. These observations have associated with then the same matrix P as encountered previously in the first method and in the first adjustment of the second method. The adjusted values of such observations are defined as

$$\tilde{\tilde{L}}^a = A_2 X_2 \quad . \tag{9.30}$$

The predicted values of this type are

$$\tilde{\tilde{L}}'^a = A_2' X_2 \quad , \tag{9.31}$$

where the prediction points coincide with those of the first method (the

same A_2' appeared in equation 9.20). Accordingly, we have

$$\sum_{\approx L^a} = A_2 \sum_{X_2} A_2^T \quad , \tag{9.32}$$

$$\sum_{\approx L'^a} = A_2' \sum_{X_2} A_2'^T \quad . \tag{9.33}$$

In agreement with our statement that $\tilde{V}$ acts as new observations having the opposite sign, we write

$$V = A_2 X_2 + \tilde{V} \quad . \tag{9.34}$$

The remarkable property of this equation is that the vector V is in fact the vector of residuals in the original simultaneous adjustment treated in the first method; this is seen immediately upon considering (9.25) and (9.16). Equations (9.30) and (9.34) indicate that the adjusted new observations can be also written as

$$\tilde{L}^a = - \tilde{V} + V \quad . \tag{9.35}$$

Our task is to show that with the weight matrix P of the original observations (see equation 9.15) and with $\tilde{V}$ computed in the first adjustment, the solution for X_2 will be the same as the one seen in the simultaneous adjustment, i.e., as equation (9.17b) of the first method. Under the least squares condition, the solution of (9.34) is

$$X_2 = - (A_2^T P A_2)^{-1} A_2^T P \tilde{V} \quad ,$$

$$\sum_{X_2} = (A_2^T P A_2)^{-1} \quad .$$

From (9.25) and the result below it, we find that

$$\tilde{V} = - A_1 (A_1^T P A_1)^{-1} A_1^T P L + L \quad .$$

But due to $A_1^T P A_2 = 0$ characterizing the present type of parameters, we obtain

$$X_2 = - (A_2^T P A_2)^{-1} A_2^T P L \quad .$$

Thus both the X_2 and its variance -covariance matrix agree with the results of the first method.

Considering (9.19) through (9.22) together with (9.26) through (9.29) and (9.30) through (9.33), it is clear that for the adjusted observations one must have

$$L^a = \tilde{L}^a + \tilde{\tilde{L}}^a$$

with similar formulas for the predictions and for the variance-covariance matrices. All these formulas will be listed below in the review of the most important expressions. The discussion will then be concluded with a schematic representation of a few quantities related to the present adjustment.

First adjustment:

$$\tilde{V} = A_1 X_1 + L \quad , \qquad L = L^0 - L^b \quad ,$$

$$X_1 = - (A_1^T P A_1)^{-1} A_1^T P L \quad , \qquad P = \sum_{L^b}^{-1} \quad ,$$

$$\sum_{X_1} = (A_1^T P A_1)^{-1} \quad ;$$

$$\tilde{L}^a = L^0 + A_1 X_1 \qquad (= L^b + \tilde{V}) \quad ,$$

$$\sum_{\tilde{L}^a} = A_1 \sum_{X_1} A_1^T \quad ,$$

$$\tilde{L}'^a = L'^0 + A_1' X_1' \quad ,$$

$$\sum_{\tilde{L}'^a} = A_1' \sum_{X_1} A_1'^T \quad .$$

Second adjustment:

$$V = A_2 X_2 + \tilde{V} \quad ,$$

$$X_2 = - (A_2^T P A_2)^{-1} A_2^T P \tilde{V} \quad ,$$

$$\sum_{X_2} = (A_2^T P A_2)^{-1} \quad ;$$

$$\tilde{\tilde{L}}^a = A_2 X_2 \quad (= - \tilde{V} + V) \quad ,$$

$$\sum_{\tilde{\tilde{L}}^a} = A_2 \sum_{X_2} A_2^T \quad ,$$

$$\tilde{\tilde{L}}'^a = A_2' X_2 \quad ,$$

$$\sum_{\tilde{\tilde{L}}'^a} = A_2' \sum_{X_2} A_2'^T \quad .$$

Final adjusted observations and predictions:

$$L^a = \tilde{L}^a + \tilde{\tilde{L}}^a \quad (= L^b + V) \quad ,$$

$$\sum_{L^a} = \sum_{\tilde{L}^a} + \sum_{\tilde{\tilde{L}}^a} \quad ;$$

$$L'^a = \tilde{L}'^a + \tilde{\tilde{L}}'^a \quad ,$$

$$\sum_{L'^a} = \sum_{\tilde{L}'^a} + \sum_{\tilde{\tilde{L}}'^a} \quad .$$

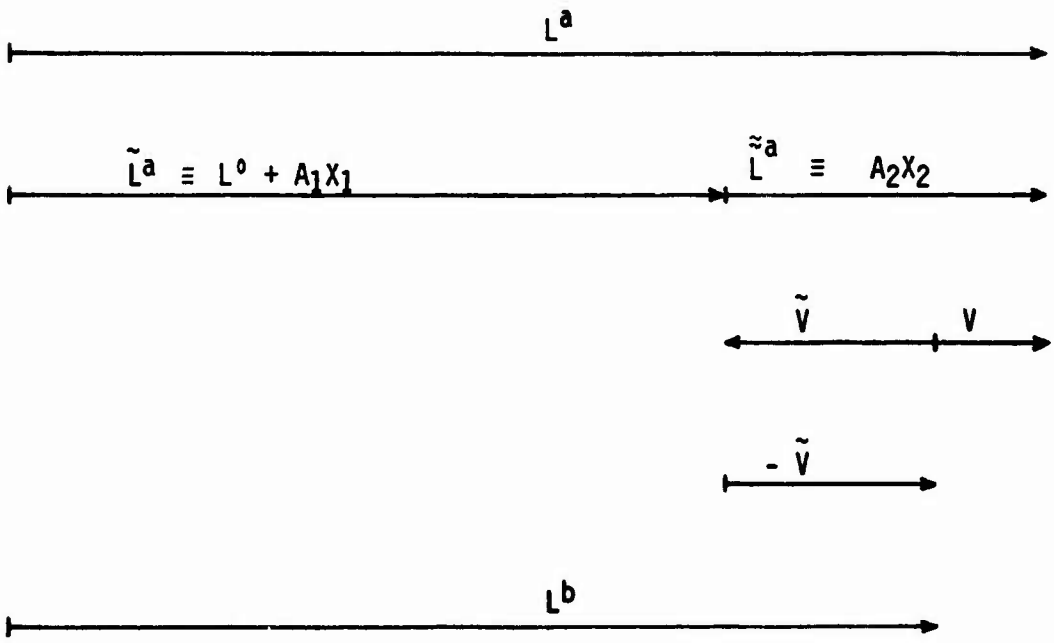


FIGURE 9.2

Schematic representation of a few quantities related to the adjustment of two independent sets of parameters

9.4.2 Application in the Point Mass Model

The adjustment of two independent sets of parameters that we have just discussed had been encountered earlier, in Section 8.1. In particular, in spite of different notations, equations (8.16), (8.17), and (8.18) of that section correspond essentially to the new equations (9.16), (9.25), and (9.34), respectively. Equations (8.16) to (8.18) dealt only with gravity anomalies. However, it is clear from the context of Section 8.1 that without any substantial modifications, they could have been applied to gecid undulations as well. An important conclusion that could be reached from these equations and from the example that followed (8.18) is that under certain ideal conditions, the spherical harmonic coefficients can be adjusted in parts, whether gravity anomalies or geoid undulations are the observables. In practice, such conditions may be fulfilled only approximately, but the adjustment performed in parts could yield acceptable results and it would always be very beneficial from the computational standpoint. This is corroborated by a quote from |Dragg, 1963|, page 34:

> "The most rational method, according to Zhongolovich, is to determine the value of harmonics gradually in the sequence of increasing degrees. In this way, the coefficients of the second degree are determined from the normal equations, and the third degree coefficients are determined from the remaining lack of agreement with observations, and so forth."

The principle of adjusting the spherical harmonic potential coefficients in two (or more) parts could be extended to adjusting the parameters in one or more sets of potential coefficients and one or more sets of point mass magnitudes. For we have seen that a relatively dense coverage by gravity anomalies or, almost equivalently, by point masses has the role of

describing finer features of the gravity field, which would be also achieved by higher degree and order potential coefficients. In fact, the resolution obtained with the equal area blocks of gravity anomalies (or point masses) with the block size $\Delta\theta$ would roughly correspond to the resolution obtained with the degree and order (k, k), where $k = 180^0/\Delta\theta^0$. Thus, for the practical purpose of this analysis, we shall consider a separate adjustment of the potential coefficients describing the long wavelength features of the gravity field, and of the point masses describing finer features.

We note that the gravity anomaly model and the geoid undulation model are linear in the potential coefficients. They could be therefore expanded around any values of the coefficients; in particular, the initial values of the potential coefficients in (8.16) to (8.18) were taken as zero because this was suitable for the theoretical derivations. However, when dealing with satellite altimetry, the mathematical model on the whole is no longer linear in the potential coefficients, and their best available values should be adopted in order to avoid unnecessary iterations in the adjustment. On the other hand, we have seen that the point mass magnitudes have no effect on the satellite position and velocity components. Furthermore, the gravity anomalies and the geoid undulations will be shown to be linear in these parameters. Thus on the whole, the mathematical model will be linear in the point mass magnitudes and their initial values will be always adopted as zero. This approach is compatible with (9.12) and (9.13) or (9.13'), where the initial values of the first group of parameters (potential coefficients) were in general nonzero, and the initial values of the second group of parameters (point mass magnitudes) were zero.

We shall now present the expressions that will introduce the point masses in the gravity anomaly model and in the geoid undulation model. The disturbing potential at the observation point (i) in terms of the point mass parameters is

$$T_i = \sum_j (1/\ell_{ij})(kM)_j \quad , \tag{9.36}$$

where j identifies the point mass, (kM) represents the point mass parameter (here the point mass magnitude is scaled by the gravitational constant), and ℓ_{ij} is the distance between the observation point i and the point mass j. In theory, the summation would extend over all the point masses, but in practice, it is limited to a certain spherical cap (for example, the cap associated with N has been represented by Ψ_o, etc.). The above equation corresponds to (2.6) of $\left|N\right|$.

In considering the boundary condition in spherical approximation, one can obtain the gravity anomaly model in terms of the point masses. It appeared in $\left|N\right|$, equation (2.19), as follows:

$$\Delta g_i = (1/R_i) \sum_j (1/\ell_{ij}) \left|(R_i^2 - F_{ij})/\ell_{ij}^2 - 2\right| (kM)_j \ , \tag{9.37}$$

where

$$R_i = (X_i^2 + Y_i^2 + Z_i^2)^{1/2}$$

is the length of the radius vector to the point i,

$$F_{ij} = X_i X_j + Y_i Y_j + Z_i Z_j$$

is the dot product between the radius vector to the point i and the radius vector to the point mass j, and

$$\ell_{ij} = \left[(X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2\right]^{\frac{1}{2}}$$

is the distance between the two points as described earlier. The geoid undulation model is represented by Bruns formula

$$N_i = T_i / G \quad ,$$

where G is the usual average value of gravity. From (9.36) we thus immediately obtain

$$N_i = (1/G) \sum_j (1/\ell_{ij})(kM)_j \quad . \tag{9.38}$$

The equations (9.37) and (9.38) have been derived for the models containing no parameters other than the point mass magnitudes. In practice, however, one group of parameters will consist of the potential coefficients. What can be said, then, about these models if the point masses are used to describe only some "left-over" effects? Will the form of equation (9.37) and (9.38) be the same if only the corresponding "left-over" point masses are included as is the case discussed presently? We shall first try to answer these questions in a heuristic manner, compatible with some previous arguments and discussions. For this purpose, a set of point masses associated with (9.37) and (9.38) could be imagined as divided into two groups. The first group would contain relatively large, deeply seated masses separated by proportionately large distances. The second group would densify this basic configuration. It would contain many more smaller masses at

shorter distances, etc., identified previously as "left-over" point masses. The two groups could be adjusted simultaneously (a theoretical case) or in two parts (the practical case that concerns us), but the model equation (9.37) or (9.38) would remain necessarily the same for either group. This follows from the possibilty to adjust separately the lower degree and order potential coefficients, and the higher degree and order potential coefficients; the assumption used in this argument is that for a global point mass coverage, the first, basic group of point masses is equivalent to the first set of potential coefficients and the second, densifying group of point masses, to the second set of coefficients. In line with this assumption is the possibility to replace the basic group of point masses by the first set of potential coefficients. In the theoretical case, we have thus arrived at a simultaneous adjustment of potential coefficients and densifying point masses. Finally, the practical case is characterized by a separate adjustment of potential coefficients followed by an adjustment of densifying point masses. The important outcome of this reasoning is that in either case, the (densifying) point masses are treated exactly as depicted in (9.37) or (9.38).

The above argument can be justified through a direct, more rigorous approach. It will also give more credibility to some assumptions and rationalizations used earlier. The main point of this approach resides in considering a higher order surface instead of an ellipsoid. Such a surface has been encountered when dealing with a basic set of spherical harmonic potential coefficients, and it is intimately connected with a higher order gravity field as compared to the normal (or ellipsoidal) field. As

in the normal field, the boundary condition and Bruns formula in the spherical approximation are expressed by

$$\Delta g = - \partial T/\partial R - (2/R)\, T \quad , \tag{9.39}$$

$$N = T / G \quad . \tag{9.40}$$

However, an important deviation from the equivalent expressions in the normal field is that the disturbing potential T now refers to the higher order field. This means that if (9.39) and (9.40) are written in terms of the point masses, one arrives again at (9.37) and (9.38), but this time, the "anomaly masses" generate the new disturbing potential of the field described by the above potential coefficients, and not the usual distrubing potential of the normal field. Consequently, these masses are nothing else but the densifying masses of the previous paragraphs, and they would clearly appear in the pertinent adjustment models together with the potential coefficients. We have thus arrived directly at the result of the discussion in the last paragraph, except for a subsequent separation of the adjustment into two parts.

The observation equations in terms of only the potential coefficients appeared in (9.25). This equation applies for either gravity anomalies or geoid undulations, and it had been expressed explicitly in the early chapters. If we wanted to consider the theoretical case, i.e., a simultaneous adjustment of the potential coefficients (X_1) and the point masses (X_2), the right-hand side of (9.25) would receive another term in agreement with (9.16). This term is symbolized by

$$dF = A_2 \, X_2 \quad . \tag{9.41}$$

In agreement with (9.37), the quantity added to one gravity anomaly observation equation would be

$$d\Delta g_i = \left[\ldots a_{ij} \ldots \right] X_2 \quad , \tag{9.42a}$$

where

$$a_{ij} = \left[(R_i^2 - F_{ij}) / \ell_{ij}^2 - 2 \right] / (R_i \ell_{ij}) \quad , \tag{9.42b}$$

$$X_2 = \begin{bmatrix} \vdots \\ (kM)_j \\ \vdots \end{bmatrix} \quad . \tag{9.43}$$

Considering (9.38) associated with N, we can similarly write

$$dN_i = \left[\ldots b_{ij} \ldots \right] X_2 \quad , \tag{9.44a}$$

where

$$b_{ij} = (1 / G)(1 / \ell_{ij}) \quad . \tag{9.44b}$$

Equation (9.42a) would be simply joined to the right-hand side of the observation equation for Δg; however, the situation with (9.44a) is slightly more complicated. The reason for this is that a satellite altimetry measurement helps to determine r_i (the radial distance to a sub-satellite point) and not N_i, and that the observation equations of satellite altimetry contain also state vector parameters. Since to a good approximation we have

$$N_i = r_i - r_i' \quad ,$$

where r_i' is the radial distance to the ellipsoid (it is measured along the same direction as r_i) which does not depend on the point mass parameters,

we can write

$$dr_i = dN_i$$

and use (9.44a). The presence of the state vector parameters does not affect this procedure because they have been shown to be independent of the point masses. The only additional remark we can make concerns the sign of (9.44a). The satellite altimetry observation equations contained $-r_i$ as well as the partial derivatives of $-r_i$, so that the final quantity to be joined to each observation equation is

$$-dr_i = \left[\ldots -b_{ij} \ldots \right] X_2 \quad . \tag{9.45}$$

Otherwise, the simultaneous adjustment of gravity anomalies and satellite altimetry would be carried out exactly in the same way as if dealing with the potential coefficients alone; each row of the "design matrix" would be simply lengthened to accommodate the new entries a_{ij} or b_{ij}, and the column vector containing the parameters would be lengthened to accommodate the elements of X_2. With this provision, the standard formulas for variance-covariance propagation, etc., would be preserved.

The description of the practical approach is equally simple. The first adjustment is performed in terms of the potential coefficients as described in the early chapters. The residuals of this adjustment are grouped in the vector $\tilde{V}$, the adjusted observations in the vector $\tilde{L}^a$, the corresponding variances-covariances in the matrix $\Sigma_{\tilde{L}^a}$, the predictions in the vector $\tilde{L}'^a$, and the variances-covariances of the predictions are grouped in the matrix $\Sigma_{\tilde{L}',a}$. The results of the second adjustment are obtained according to the summary of formulas listed at the end of Section 9.4.1. We present some of these formulas adapted to the present context:

$$v_i(\Delta g) = \left[\cdots a_{ij} \cdots \right] X_2 + \tilde{v}_i(\Delta g),$$

$$v_i(\text{altim.}) = \left[\cdots - b_{ij} \cdots \right] X_2 + \tilde{v}_i(\text{altim.}) \quad ;$$

$$\tilde{l}_i^a(\Delta g) = \left[\cdots a_{ij} \cdots \right] X_2 \quad ,$$

$$\tilde{l}_i^a(\text{altim.}) = \left[\cdots -b_{ij} \cdots \right] X_2 \quad ;$$

$$\tilde{\tilde{\sigma}}_i^2(\Delta g) = \left[\cdots a_{ij} \cdots \right] \sum_{X_2} \begin{bmatrix} \vdots \\ a_{ij} \\ \vdots \end{bmatrix} ,$$

$$\tilde{\tilde{\sigma}}_i^2(\text{altim.}) = \left[\cdots - b_{ij} \cdots \right] \sum_{X_2} \begin{bmatrix} \vdots \\ -b_{ij} \\ \vdots \end{bmatrix} \quad ;$$

$$\tilde{\tilde{l}}_i'^a(\Delta g) = \left[\cdots a'_{ij} \cdots \right] X_2 \quad ,$$

$$\tilde{\tilde{l}}_i'^a(N) = \left[\cdots + b'_{ij} \cdots \right] X_2 \quad ;$$

$$\tilde{\tilde{\sigma}}_i'^2(\Delta g) = \left[\cdots a'_{ij} \cdots \right] \sum_{X_2} \begin{bmatrix} \vdots \\ a'_{ij} \\ \vdots \end{bmatrix} ,$$

$$\tilde{\tilde{\sigma}}_i'^2(N) = \left[\cdots + b'_{ij} \cdots \right] \sum_{X_2} \begin{bmatrix} \vdots \\ +b'_{ij} \\ \vdots \end{bmatrix}$$

The same summary of formulas leads also to the following final results:

$$l_i^a(\Delta g) = \tilde{l}_i^a(\Delta g) + \tilde{\tilde{l}}_i^a(\Delta g) \quad ,$$

$$l_i^a(\text{altim.}) = \tilde{l}_i^a(\text{altim.}) + \tilde{\tilde{l}}_i^a(\text{altim.}) \quad ;$$

$$\sigma_i^2(\Delta g) = \tilde{\sigma}_i^2(\Delta g) + \tilde{\tilde{\sigma}}_i^2(\Delta g) \quad ,$$

$$\sigma_i^2(\text{altim.}) = \tilde{\sigma}_i^2(\text{altim.}) + \tilde{\tilde{\sigma}}_i^2(\text{altim.}) \quad ;$$

$$l_i'^a(\Delta g) = \tilde{l}_i'^a(\Delta g) + \tilde{\tilde{l}}_i'^a(\Delta g) \quad ,$$

$$l_i'^a(N) = \tilde{l}_i'^a(N) + \tilde{\tilde{l}}_i'^a(N) \quad ;$$

$$\sigma_i'^2(\Delta g) = \tilde{\sigma}_i'^2(\Delta g) + \tilde{\tilde{\sigma}}_i'^2(\Delta g) \quad ,$$

$$\sigma_i'^2(N) = \tilde{\sigma}_i'^2(N) + \tilde{\tilde{\sigma}}_i'^2(N) \quad .$$

It is emphasized that the above formulas apply only after appropriate constraints have been included in the adjustment. The constraints are the subject matter of the next section. In practice, they may have the form of observation equations, assigned very large weights. In the theoretical (simultaneous) adjustment, the constraints would involve only the parameters in the second set (point mass magnitudes). In the practical approach, these constraints would be

included only in the second adjustment. Although they are not believed to introduce significant changes in unconstrained residuals, the local constraints should always be included in the adjustment of point masses; by constrast, the group of global constraints appears to be superfluous, if harmless.

We may add one last practical remark regarding the error propagation. We have seen that

$$\Sigma_{L^a} = \Sigma_{\tilde{L}^a} + \Sigma_{\tilde{\tilde{L}}^a} ,$$

with a similar formula holding true for variances-covariances of the predictions. The first adjustment in terms of potential coefficients yields $\tilde{L}^a$ and $\Sigma_{\tilde{L}^a}$, which would be expected to approach true values only if all the remaining potential coefficients were known to be perfect zeros. Otherwise $\tilde{L}^a$ contains biases due to unmodeled parameters and $\Sigma_{\tilde{L}^a}$ is unrealistically optimistic for the same reason. We may then conclude that the above equation should not be interpreted in the sense that the actual uncertainty in the results L^a increases with every added parameter (here with added point masses), merely because the nominal variances get larger. Rather, as more and more appropriate parameters are introduced into the adjustment, the elements in Σ_{L^a}, etc., approach better and better the true variances-covariances.

9.5 Constraints

9.5.1 Local Constraints

After an adjustment of spherical harmonic potential coefficients and point mass parameters, a function (F^a) may be written as composed of two parts, namely

$$F^a = F_{SH} + F_{PM} ;$$

in this expression, SH indicates the part generated by the potential coefficients and PM, by the point masses. In an area whose size is not too small with respect to the resolution possibilities, we should reasonably expect that

$$(1 / \Delta A) \int_{\Delta A} (F_{SH} + F_{PM}) \, dA \simeq \bar{F}_{(\Delta A)true} , \qquad (9.46)$$

where dA is the element of ΔA and where the right-hand side represents the true mean of the function F in the area ΔA.

We shall be particularly interested in finding surface regions ΔA of such sizes that

$$(1 / \Delta A) \int_{\Delta A} F_{SH} \, dA \simeq \bar{F}_{(\Delta A)\,true} \qquad (9.47)$$

or

$$\int_{\Delta A} F_{SH} \, dA \simeq \Delta A \, \bar{F}_{(\Delta A)\,true} \qquad (9.47')$$

can be considered fulfilled. For if (9.47) is valid, it gives rise to a

constraint of the type

$$(1 / \Delta A) \int_{\Delta A} F_{PM} \, dA \simeq 0 \quad , \tag{9.48}$$

as may be gathered upon consulting (9.46). The departure of the left-hand side minus the right-hand side of (9.47) from zero will be called here the "mean deviation". The term "deviation" will describe such a departure in connection with (9.47'). A simple equivalent of (9.48) for practical applications is

$$\int_{\Delta A} F_{PM} \, dA \simeq 0 \quad . \tag{9.48'}$$

In order to assess if and when an equation like (9.47) applies, we have to examine the relationship between the "SH" part of a function and the true function. As a first step in this direction, we borrow a statement from |N|, page 73:

> "If the degree and order are limited to (14, 14) for example, the described fields are heavily smoothed representations of the true fields. As such, one would expect the mean value of the true function and the mean value of the smoothed function over an area to approach a common value as the size of the area is increased."

If we want to arrive at a set of useful constraints, we have to provide sensible answers to the following questions. What is the smallest area ΔA over which (9.47) can still be satisfied? Does ΔA correspond to the shortest half wavelength of the spherical harmonic resolution as it is assumed in |N|, or does it correspond to an intuitively more appealing whole wavelength? And what kind of function is likely to give the best

results in conjunction with a given ΔA? We shall seek answers to these questions by means of heuristic reasoning, since we are dealing with some approximate relations and desirable properties of a very general nature. We shall be assisted by a few simple one-dimensional examples. In these examples, $\Delta\theta$ will indicate the shortest half wavelength which, in the spherical harmonic model symbolized by (k, k), was seen to be obtained from the approximate relation $\Delta\theta^0 \simeq 180^0 / k$.

We start with one of the simplest possible cases, pictured in Figure 9.3. It is assumed that the true function consists of merely one type of wave and that many redundant observations to determine its shape have been made. Ideally, the (one-dimensional) "SH function" determined from these observations in a least squares adjustment would coincide with the true function. On the other hand, the true means within the regions $\Delta\theta$ associated with the individual mid-regions are denoted by dots in Figure 9.3. In this example, no deviations defined following equation (9.48) exist. The relations (9.47) or (9.47') are thus fulfilled exactly in arbitrary regions ($\Delta\theta$, $2\Delta\theta$, $0.1\Delta\theta$, etc.).

Consider next the same true function, but significantly fewer observations. In fact, let the observations be represented by (ideally) true means in the regions $\Delta\theta$, in much the same way as gravity anomalies would be represented by the mean values in some equal area blocks. Suppose further that there are no redundant observations, similar to the two-dimensional case with the number of blocks just sufficient to determine all the potential coefficients through the degree and order (k, k); we have seen that such blocks would have the sides approximately equal to π/k. As depicted

in Figure 9.4, the unique solution makes the SH function pass through the dots, described in the last paragraph. It appears that the true function and the smoothed SH are now different. This situation can be related to an earlier statement (it followed 9.7b) that one observation per shortest half wavelength is not, in general, sufficient for a good determination of the corresponding fine features.

However, an important point to be made in this connection is that (9.47) or (9.47') are in general not fulfilled in regions $\Delta\theta$. Let us first consider the "original regions" $\Delta\theta$ delimited by the vertical lines and identified by the letters A, B, C, and D. The deviations taken as the integrals of $F_{SH} - F_{true}$ over these regions are very large. It is more practical to visualize them in terms of a signed difference between the area below (or above) the curve within one such region and the area of the corresponding rectangle. These deviations are described above the letters A, B, C, D. Their magnitude is symbolically represented as varying from almost zero to large, according to the convention:

0, (+) or (-), + or -, ++ or --,

the last group meaning "very large positive" or "very large negative". However, the regions $\Delta\theta$ could be chosen in many different ways, yielding quite different deviations. One useful set of such new regions, called "shifted regions" $\Delta\theta$, has been indicated in Figure 9.4 by the vertical dashed lines. They bear the identification letters a, b, c, and d. The true means in these regions coincide with the abscissa and all the deviations are now zero (see "0" written underneath each of a, b, c, d). The borderlines of the shifted regions $\Delta\theta$ coincide with the central lines of the original regions $\Delta\theta$. A similar convention will be observed in the following two figures.

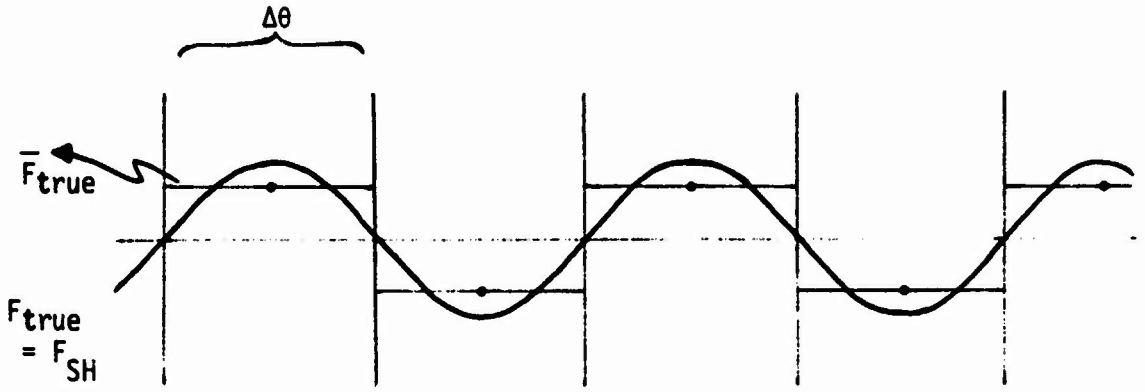


FIGURE 9.3

The simplest case. Good redundancy, F_{SH} coincides with F_{true}; no deviations exist.

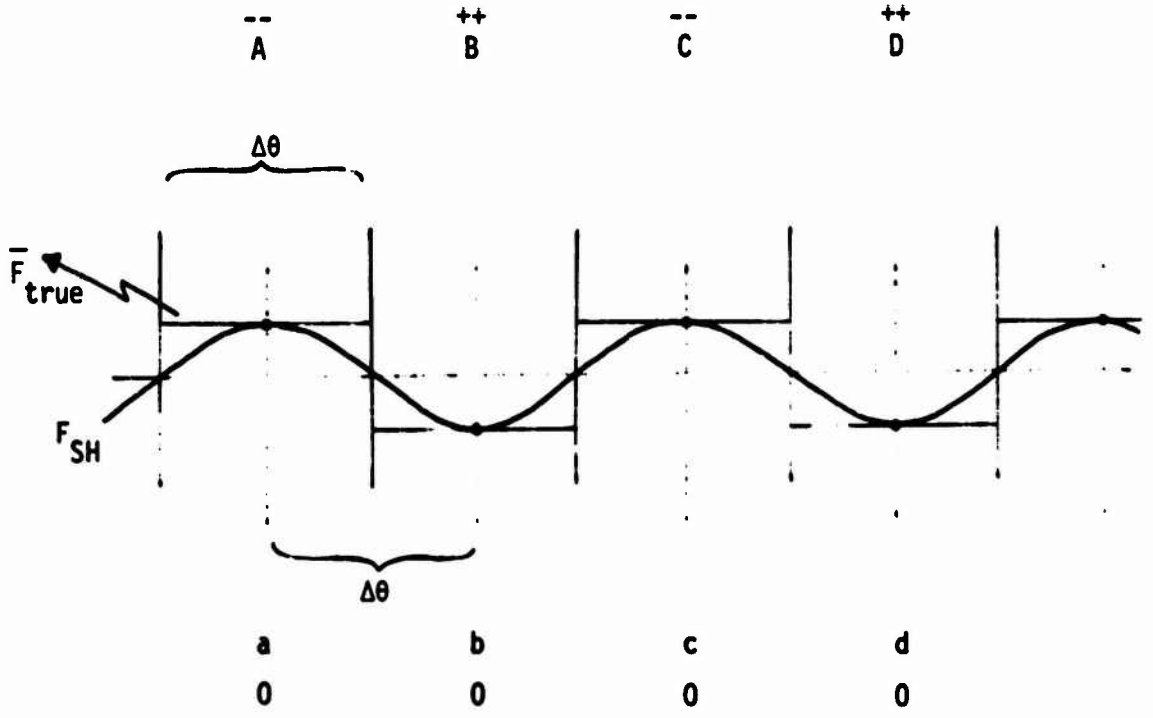


FIGURE 9.4

The simplest case. No redundancy, F_{SH} coincides with $\bar{F}_{true}$ at the central points; deviations are large in original regions Δθ, zero in shifted regions Δθ. No deviations would exist in 2Δθ regions (original or shifted).

In this example, (9.47) or (9.47') are not satisfactorily fulfilled when considering the regions $\Delta\theta$ (these regions correspond to the shortest half wavelength). On the other hand, if the regions $2\Delta\theta$ are adopted, the deviations would become zero no matter where the vertical lines were placed. Considering the solid vertical lines in Figure 9.4, we observe that the resulting discrepancy is zero between the first and the third line, between the second and the fourth line, etc. This of course agrees with the symbols -- and ++ associated with A and B, ++ and -- associated with B and C, etc. We would similarly have zero discrepancy between the first and the third dashed line, etc. This simplified example thus indicates that the regions $2\Delta\theta$ may be the smallest to have the desired properties, even in the case when few or no redundant observations exist. In terms of the blocks on the sphere, the size $2\Delta\theta \times 2\Delta\theta$ appears thus far to be a good choice for the "constraint blocks" (the spherical angular distance accross the block is of course greater than $2\Delta\theta$, but some circular areas ΔA would be far more complicated and they would offer no substantial advantages).

If half wavelengths longer than $\Delta\theta$ were present, the outcome of the simplified example would not be altered to any great extent. The reason for this is that longer wavelengths would act as a "trend" which is likely to be reflected by F_{SH} better than the shorter wavelengths could be. In fact, the situation of Figure 9.4 would lead to redundancy, as far as the half wavelengths longer than $\Delta\theta$ are concerned. The half wavelengths that are shorter than $\frac{1}{2}\Delta\theta$ are not a cause for concern, either. Under the assumption that the true function is reasonably smooth -- and this assumption will be crucial in selecting the type of function that should serve for constraints -- the deviations are likely to be very small. There will be at least two

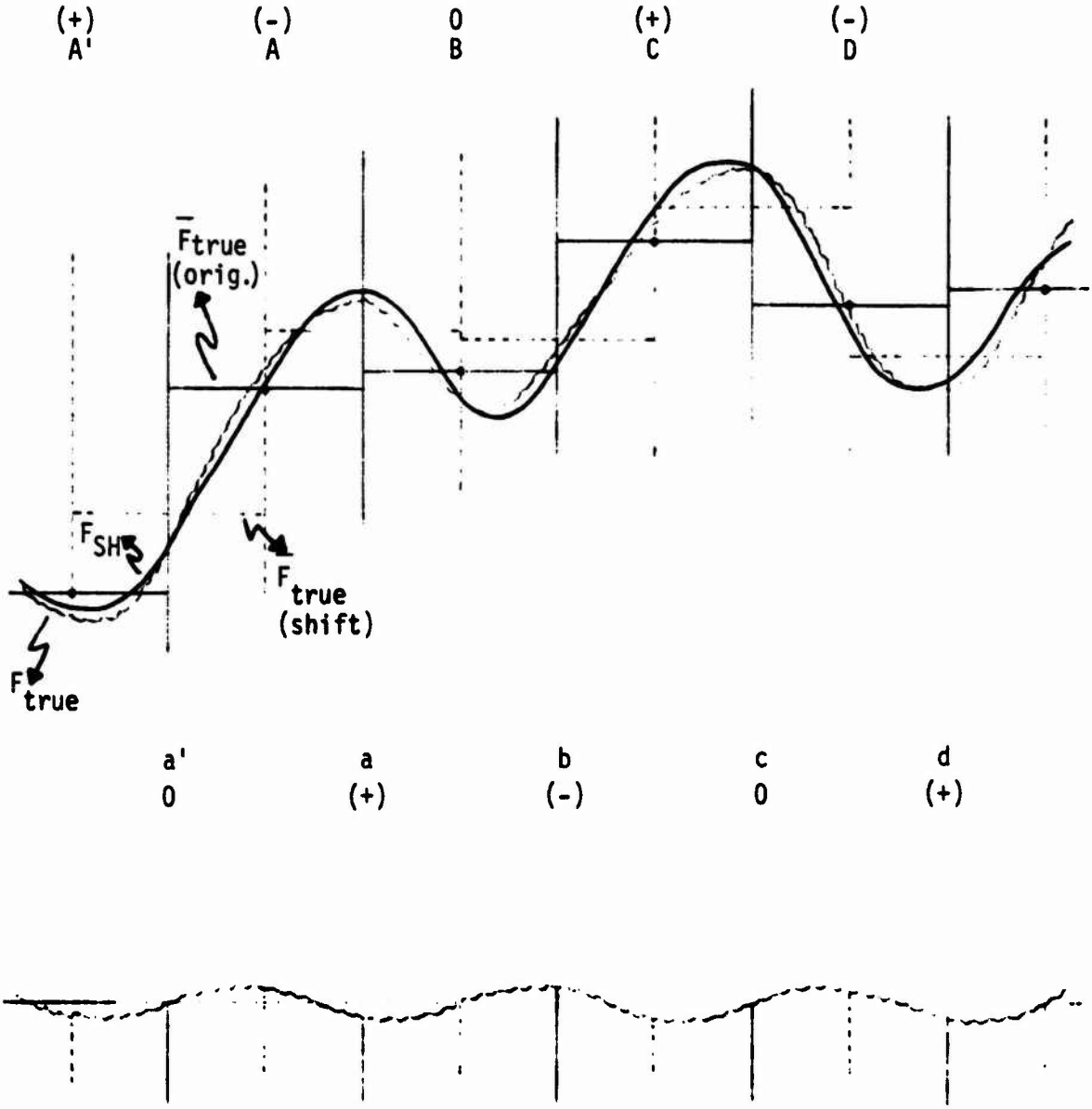


FIGURE 9.5

The more realistic case. Good redundancy, F_{SH} is close to F_{true} (differences are seen better on the lower picture); deviations are small in original or shifted regions $\Delta\theta$. By comparison, deviations -- and even more the mean deviations -- would in general decrease significantly in $2\Delta\theta$ regions (original or shifted).

full waves (or four half-waves) in any region $2\Delta\theta$ and the integrated differences will tend to cancel out. The situation which should be examined more closely comprises the half wavelengths between $\frac{1}{2}\Delta\theta$ and $\Delta\theta$. This will be done in the next example, in which also some long wavelengths will be present.

A more realistic example than the one just discussed appears in Figure 9.5, also containing the most important verbal explanations. Several features of this figure (original regions $\Delta\theta$, shifted regions $\Delta\theta$, their identifications, etc.) resemble those of Figure 9.4. Many redundant observations are now present, which explains why F_{SH} does not pass through the mean values at the original mid-regions $\Delta\theta$. The agreement between F_{SH} and F_{true} is therefore reasonably good, both in the original and in the shifted regions $\Delta\theta$. The existing deviations in the individual regions $\Delta\theta$ are attributed, in this example, mostly to the features whose half wavelength is $0.75\ \Delta\theta$. However, under the assumption of a smooth function an important improvement occurs if the deviations are considered in the regions $2\Delta\theta$ instead. The improvement is even more noticeable for the mean deviations (the "regular" deviations decrease, while at the same time the size of the individual regions increases). In such a case, many cancellations between positive and negative deviations occur, as can be observed, for example, in the regions C, D or a, b, etc.

We shall finally consider the same example as above, but with no redundancy. This case, together with the most important verbal explanations is presented in Figure 9.6. The curve representing F_{SH} now passes through the dots associated with $\bar{F}_{true}$ in the original regions $\Delta\theta$, just as it did

in Figure 9.4. The deviations in the original or shifted regions $\Delta\theta$ are in general very large, as can be observed in A', A, B, etc. However, dramatic improvements would occur if $2\Delta\theta$ regions, original or shifted, were considered instead (the regions A', A, or A, B, etc., would be joined together). A similar statement could be made with regard to the shifted regions a, b, etc.

There are several conclusions which can be drawn even from such simple examples as we have just seen. First of all, the proposition that the constraint blocks

$$\Delta A = 2\Delta\theta \times 2\Delta\theta \tag{9.49}$$

are more appropriate than the original blocks

$$\Delta S = \Delta\theta \times \Delta\theta \tag{9.49'}$$

emerges as reasonable. The relations (9.47) or (9.47') appear to hold much more satisfactorily with ΔA than with ΔS. This is true not only if no or almost no redundancy exists, but also in practical cases that concern us, where many redundant observations may be available. It should be mentioned that in $|N|$, the possiblity of using larger constraint blocks than those depicted in (9.49') has not been entirely discarded. This is recognized upon reading the statement on page 75:

> "A conservative approach would be to apply the condition only to regions that were considerably greater than the area given by equation (5.2)."

The just cited equation (5.2) corresponds roughly to our equation (9.49').

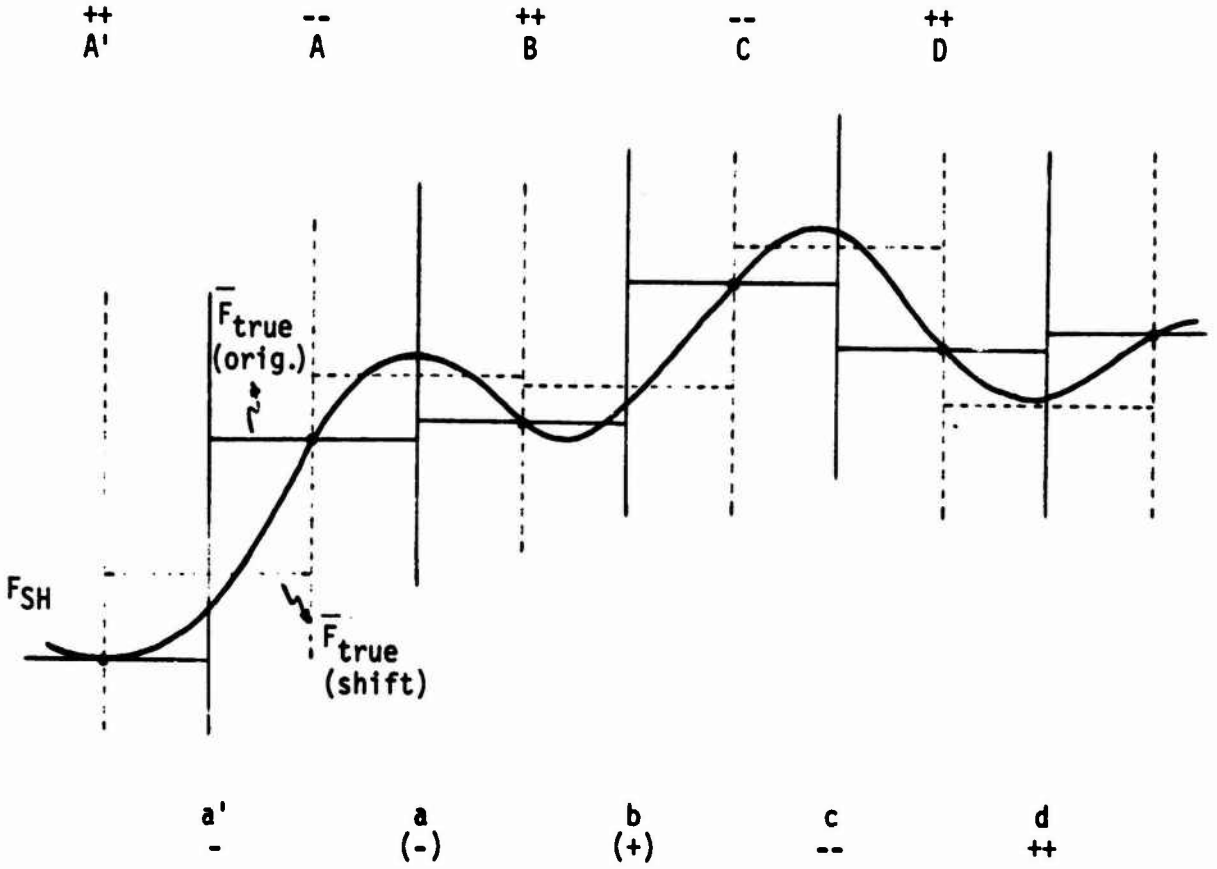


FIGURE 9.6

The more realistic case. No redundancy, F_{SH} coincides with $\bar{F}_{true}$ at the central points; deviations are generally large in original or shifted regions $\Delta\theta$. Drastic improvements would occur in $2\Delta\theta$ regions (original or shifted).

Another important conclusion we have reached is that the function to which (9.47) or (9.47') may be applied should be smooth. If (9.2b) and (9.3) are considered, it becomes clear that a likely candidate for such constraints is N and not Δg. Although the wavelength is the same in both cases, the function represented by Δg is much rougher than the one represented by N, due to the factor (n - 1) in the former. Consequently, (9.47) and (9.48') will now be specialized for geoid undulations as follows:

$$(1 / \Delta A) \int_{\Delta A} N_{SH} \, dA \simeq \bar{N}_{(\Delta A)true} \quad ,$$

$$\int_{\Delta A} N_{PM} \, dA \simeq 0 \quad . \tag{9.50}$$

We notice that in $|N|$, a constraint of the type (9.48') has not been applied to gravity anomalies, either.

In analogy to the two kinds of constraints used in $|N|$, we stipulate that the total volume and the total mass in the volume regions subtended by individual constraint blocks ΔA should be unchanged by the presence of the point masses. The first stipulation is fulfilled by (9.50) and the second, by a constraint of the type

$$\sum_j (kM)_j = 0 \quad , \tag{9.51}$$

where j ranges over all the point masses within the block ΔA. The importance of the mass magnitude constraints had been discussed in connection with the task of preserving the coordinate system. We have thus far arrived at two constraints per block ΔA. By comparison, the number of constraints in $|N|$

is two per block ΔS (one of the type 9.50, and one of the type 9.51, with ΔA being replaced by ΔS). This means that according to $|N|$, there would be eight constraints per block ΔA, since from (9.49) and (9.49') we find that

$$\Delta A = 4\Delta S \quad .$$

However, we shall shortly present an argument which will lead to an application of both kinds of constraints (i.e., 9.50 and 9.51) to overlapping blocks of approximately the same size ΔA. Accordingly, the number of constraints suggested in this analysis will be about one half -- rather than one fourth -- of their number in $|N|$.

We shall now concentrate our attention on the development of equation (9.50) for a practical application. Due to the relation $N \approx T/G$ encountered e.g. following (9.4), instead of (9.50) we can use

$$\int_{\Delta A} T_{PM} \, dA \approx 0 \quad .$$

Upon consulting (9.36), this can be written as

$$\int_{\Delta A} \sum_{j} (1 / \ell_{ij}) \, (kM)_j \, dA_i \approx 0 \quad ,$$

where dA_i is the element of the constraint block ΔA under consideration; each point i refers to the center of this element and each point j refers to a point mass. The extent to which the point masses should be included in the above summation with respect to j depends on the truncation cap (Ψ_o) around each point i. In some cases, it might even prove simpler to include

all the point masses indiscriminately, although the contribution of some of them could be completely negligible. In the form of an actual constraint, the last relation can be formulated as follows:

$$\sum_i \sum_j \; (1 / \ell_{ij}) (k M)_j \; dA_i \; = \; 0 \quad , \qquad (9.52)$$

where i again extends over all the elements in a particular ΔA.

In order to express (9.52) even more explicitly, we consider the constraint block ΔA subdivided into several geographic compartments. This is achieved by dividing its span in latitude into equal parts $d\phi$, and its span in longitude into equal parts $d\lambda$. For example, ΔA could comprise 400 elements dA_i in 20 "rows" and 20 "columns". The areas of these elements would not be all equal (they would be equal only within each "row") and they would be expressed as

$$dA_i \; = \; \cos \phi_i \; \times \; (d\phi \; d\lambda) \quad . \qquad (9.53)$$

Care should be taken to compute $\cos \phi_i$ only once for each "row". In fact, the value of $\cos \phi_i$ would not in general change by large amounts between neighboring rows. Upon substituting (9.53) into (9.52), exchanging the summation symbols, and dividing the whole equation by the nonzero quantity $(d\phi \; d\lambda)$, the constraint (9.50) may finally be written in the following, computationaly manageable form:

$$\sum_j \left[\sum_i \; (1 / \ell_{ij}) \; \cos \phi_i \right] (kM)_j \; = \; 0 \quad . \qquad (9.54)$$

As stated earlier, certain points j may be omitted, depending on the angle Ψ_o.

In the last step, overlapping regions of comparable size (ΔA) will be introduced. In order to assure a smooth transition between blocks, the same type of constraints as (9.51) and (9.54) will be applied also to these new "overlapping constraint blocks". Otherwise, discontinuities in N and in point mass magnitudes might occur near the boundaries of the original blocks. A simplified example is shown in Figure 9.7, depicting a region near the equator. The original constraint blocks are delimited by full lines and the overlapping constraint blocks, by dashed lines. The corners of the latter blocks coincide with the centers of the former blocks. In more extended regions, such an ideal situation would of course change with appreciable changes in latitude. In Figure 9.7, the central points are represented by dots, two of which are labeded as A, E. The smaller squares having the spherical angle $\Delta\theta$ on the sides correspond to the regions ΔS described earlier. One such region is indicated by the letters A, B, C, D.

We conclude this discussion by summarizing its most important results. One may first construct the appropriate original constraint blocks that have the central angle of approximately $2\Delta\theta$ on the side; $\Delta\theta$ corresponds to the shortest half wavelength of the spherical harmonic resolution. One kind of constraints (pertaining to mass) in these blocks is given by (9.51) and another kind (pertaining to volume), by (9.54). One may then proceed to construct the overlapping constraint blocks of about the same size (ΔA) as the previous blocks. The same two types of constraints are applied in conjunction with the new blocks; this about doubles the number of constraints when compared to the number of the original constraints alone. On the average, there are thus approximately four constraints per region ΔA, which corresponds to one constraint per region ΔS. By comparison, the number of constraints

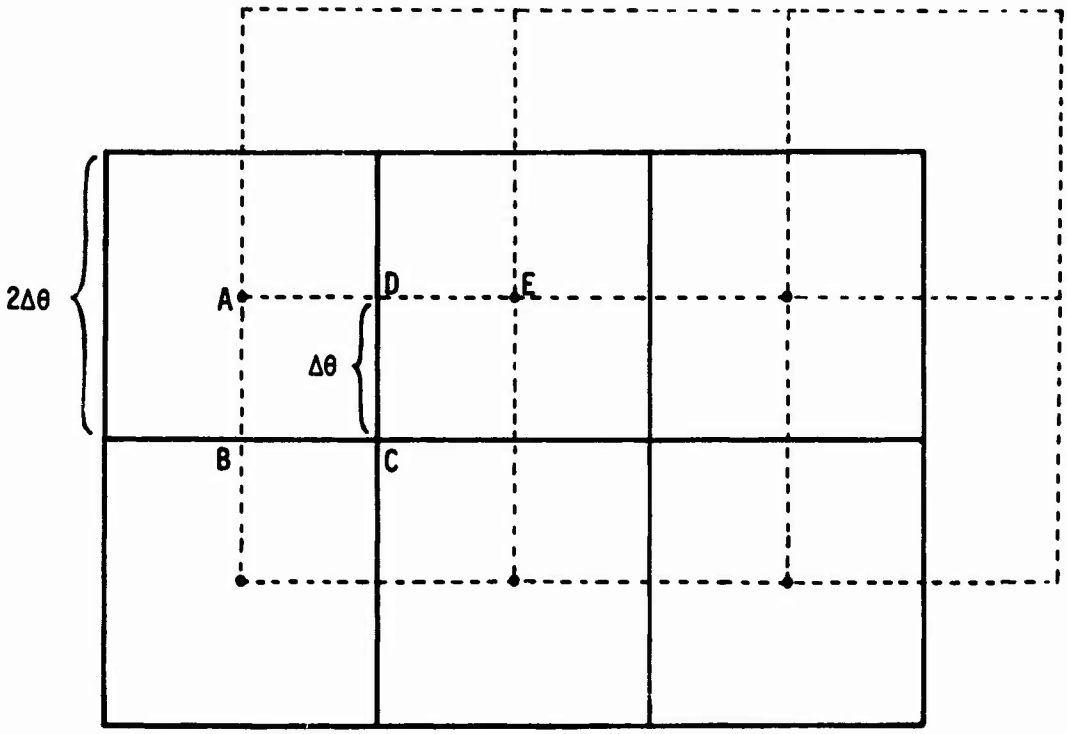


FIGURE 9.7

A limited, special example of a suggested arrangement of original constraint blocks and overlapping constraint blocks, all having the spherical angle 2Δθ on the side

used in $\lceil N \rceil$ per region ΔS was two (one pertaining to mass and the other, to volume). Althcugh there is a difference in the number of constraints between this analysis and $|N|$, the form of these constraints is similar. We close this discussion with a practical remark. The constraints (9.51) are meaningful, clearly, only in the blocks (ΔA) containing point masses. On the other hand, the constraints (9.54) should ideally comprise some additional, peripheral blocks.

9.5.2 Global Constraints

We have shown earlier that the type of constraints called "global constraints" do not have to be present in conjunction with the point masses, provided the local constraints are properly handled. The properties of the global constraints as well as their enforcement in a spherical harmonic model have been amply discussed in Section 9.1. We shall nevertheless present these constraints for the sake of completeness and for reference purposes. First, described below, will be a desired property of the coordinate system. This will be followed by a presentation of constraints needed to enforce this property in a spherical harmonic model. Finally, the explicit formulas (i.e., the global constraints) will be given that, in theory, preserve this property after the point masses have been introduced.

Property: The Cartesian origin should coincide with the earth's center of gravity.

Constraints (3) in the spherical harmonic model:

$$C_{10} = 0 \quad , \quad C_{11} = 0 \quad , \quad S_{11} = 0 \quad . \qquad (9.55a)$$

Constraints (3) in the point mass model, corresponding to (9.55a) in the same order (they can be deduced from equations 2 - 44b in |HM|):

$$\sum_j z_j\,(kM)_j = 0\ ,\quad \sum_j x_j\,(kM)_j = 0\ ,\quad \sum_j y_j\,(kM)_j = 0\ , \tag{9.55b}$$

where j ranges over all the existing point masses.

Property: The Cartesian z - axis should coincide with the earth's axis of rotation.

Constraints (2) in the spherical harmonic model:

$$C_{21} = 0\ ,\quad S_{21} = 0\ . \tag{9.56a}$$

Constraints (2) in the point mass model, corresponding to (9.56a) in the same order (they can be deduced from two of the equations 2 - 44c in |HM|):

$$\sum_j x_j z_j\,(kM)_j = 0\ ,\quad \sum_j y_j z_j\,(kM)_j = 0\ , \tag{9.56b}$$

where j ranges over all the existing point masses.

Property: The Cartesian x, y axes should be fixed with respect to the principle axes of inertia.

Constraint (1) in the spherical harmonic model:

$$S_{22} = \text{constant}. \tag{9.57a}$$

Constraint (1) in the point mass model that would preserve this configuration (it can be deduced from one of the equations 2 - 44c in |HM|):

$$\sum_j x_j y_j\,(kM)_j = 0\ , \tag{9.57b}$$

where j ranges over all the existing point masses.

Total number of constraints that can define -- or preserve -- the coordinate system: 6.

Let us finally consider the expressions

$$\Delta g_0 = (1/4\pi) \iint_\sigma \Delta g \, d\sigma \ ,$$

and

$$N_0 = (1/4\pi) \iint_\sigma N \, d\sigma \ .$$

The first of these two expressions has been further developed prior to equation (8.3b) and the second, following equation (8.4a). We have also seen in this context that if the best available values for the earth's mass and for the potential on the geoid are accepted, i.e., if it may be assumed that

$$k\,\delta M \equiv 0 \ , \qquad \delta W \equiv 0 \ , \tag{9,58}$$

then both Δg_0 and N_0 are expected to be zero (or nearly zero in practice). Such an assumption appeared prior to (8.5). A sensible approach, when dealing with a spherical harmonic adjustment model in practice, would be to adopt the above mentioned best available values and to hold them fixed. This has been indeed assumed throughout most of the present study. It is therefore expected that after the adjustment, the spherical harmonic model will yield

$$\iint_\sigma \Delta g \, d\sigma = 0 \ , \tag{9.59}$$

$$\iint_\sigma N \, d\sigma = 0 \ . \tag{9.60}$$

In using the expressions (9.39) and (9.40) for the derivation of the point mass model, one implicitly assumes that $\delta W \equiv 0$ holds true, i.e., that the point mass solution will not change the potential of the geoid. This follows from the more general relations (2 - 178') and (2 - 179') of [HM], which contain also δW. On the other hand, we have seen that the point masses are not allowed to change the total mass of the earth either, since the mass is kept unchanged already at the level of each individual local region. Consequently, (9.58) still holds true and (9.59), (9.60) are fulfilled also after the point masses have been added to the spherical harmonic adjustment model. We can therefore write for the parts of Δg and N generated by the point masses:

$$\iint_{\sigma} \Delta g_{PM} \, d\sigma = 0 \quad , \tag{9.61}$$

$$\iint_{\sigma} N_{PM} \, d\sigma = 0 \quad . \tag{9.62}$$

A local equivalent of (9.62) has been treated in Section 9.5.1. It was considered that such an expression would be satisfied reasonably well also on the global scale. In other words, equation (9.62) has been anticipated, although for other reasons than those discussed presently. However, no relation of the type (9.61) has been mentioned prior to the last paragraph. We have thus newly established that also (9.61) should hold true, without the need for any global constraints.

10. SUMMARY AND CONCLUSIONS

One of the most important objectives accomplished during the present analysis has been the upgrading of the AFGL computer program SAGG (Satellite Altimetry and Ground Gravity). This program serves in adjusting satellite altimetry data and gravity anomaly data, separately or in combination. The adjustment model is conceived in terms of spherical harmonic potential coefficients, and the adjustment is effectuated in the earth fixed coordinate system.

The highlight of the satellite altimetry approach in SAGG is the use of the short arc adjustment mode. This feature makes a determination of the geoid surface possible without the requirement of highly precise reference orbits. Approximate values of the state vector components with some fairly large a-priori standard errors are in fact treated as a set of independent weighted parameters, to be adjusted simultaneously with the potential coefficients. The observations gathered over the oceanic regions represent the distances from a satellite to the geoid. For many purposes, the geoid surface is assumed to coincide with the surface of oceans.

An original version of SAGG was described in |Blaha, 1975|. It was intended for error analysis using simulated data. Since then, many ways to improve its theoretical basis as well as its efficiency have been found and implemented. They will be now listed and briefly explained.

In the original version, the radial distance to a satellite point was differentiated with respect to the state vector parameters and the radial distance from the coordinate origin to the geoid (r) was differentiated with respect to the earth potential coefficients. The observed

satellite altimetry value (H) was considered to be equal to the difference between these two radial distances. In the present study, a correction has been introduced that makes it possible to express the mathematical model for H as accurately as practicable, good to a few centimeters. With regard to the partial differentiation, it has been shown that r, in addition to being differentiated with respect to the potential coefficients, should be differentiated also with respect to the state vector components. This gives rise to a second type of correction. To express these corrections, the ellipsoidal approximation to the geoid has been used. The results indicate that actual computation of these corrections is a very simple matter; an eventual upgrading of SAGG or any similar satellite altimetry computer program can be accomplished with almost no additional effort. A practical benefit of this refinement is a faster convergence in the adjustment, which has removed the need for iterated solutions.

Originally, adjustments were made with unrealistically large a-priori standard errors of the state vector parameters, in the mode called "coarse error spheres"; the three position sigmas were equal and the three velocity sigmas were equal as well. The real error ellipsoids related to the state vectors were well within the boundaries of these spheres. However, it is well known that in reality the error ellipsoids are far from having a spherical shape. In fact, they are elongated in the in-track component, the uncertainty being smaller in the cross-track component and the smallest in the third orthogonal component coinciding roughly with the "up" direction. Consequently, a weighting scheme has been designed and implemented into SAGG that takes into account the realistic error characteristics of the orbital information. It results in a more rigorous adjustment model and, usually, in slightly smaller a-posteriori standard errors.

In the previous version of SAGG, the earth rotation was included neither in the satellite altimetry part nor in the gravity part, in which the observable quantity was represented by a slightly modified radial component of gravity. Both mathematical models had been verified with regard to the spherical harmonic potential coefficients. Used in these tests was the method of "parameter shifts", described e.g. on page 75 of |Blaha, 1975|. Since then, the earth's rotational velocity has been introduced in the formulas of satellite altimetry. On the other hand, the adjustment of the radial component of gravity has been replaced by an equally adequate, but simpler and more efficient gravity anomaly model. This model has been deemed sufficiently accurate after a detailed study, in which the errors and approximations entailed have been traced down and analysed. In the newly introduced gravity anomaly part of SAGG, however, the rotational velocity is absent from observation equations; it cancels out upon subtracting the normal gravity from the actual gravity, since the reference ellipsoid is assumed to rotate at the same speed as the actual earth. After the introduction of these changes in SAGG, the part of the models corresponding to the potential coefficients has again been tested by the method of "parameter shifts".

In an additional effort, the orbital part of SAGG has been tested through a completely independent approach, using the method called the "analytic F and G series evaluation". This method is valid only for Keplerian orbits. The comparisons have thus been achieved by setting all the potential coefficients in SAGG equal to zero. Tests have been made with the orbits varying in inclination and eccentricity. All the final results between the two approaches (the orbital integrator in SAGG versus the analytic series) have

agreed to all the significant digits printed. As a by-product of various analyses of these tests, further improvements in SAGG have been realized. In particular, numerical inversions of certain (6 x 6) transformation matrices have been replaced by very simple analytic expressions. These matrices are used once per arc and they serve in transforming the state vector components from one coordinate system to another.

In the original version of the orbital part of SAGG, the data were simulated by an orbital integrator routine, generating events on a specified orbit which was then broken into a number of short arcs. After the recent modifications, the orbital information can also be read from a tape containing the epoch time, the state vector components in a chosen inertial coordinate system, and event times for each arc; there is no limitation on the number of arcs. It is advantageous to choose an inertial system (different for each arc) to coincide with the earth fixed system of adjustment. However, an additional, separate program development has been made in which one common inertial system may be used. This may prove useful if, in the future, the state vector components should be given in one common inertial coordinate system referred e.g. to the epoch of 1950.0. This option, too, has been thoroughly tested in SAGG.

The following improvement has been implemented in both the satellite altimetry and the gravity anomaly part of SAGG, and it applies equally well in a combined adjustment. In the original version, all the potential coefficients were either completely free to adjust, or were all weighted considering their standard error equal to their absolute value

itself (an inefficient option rarely exercised). From the practical standpoint, four tasks have been accomplished:

(a) the coefficients have been optionally weighted up to a desired degree and order,

(b) an option has been provided for weighting of only the zonal coefficients up to a desired degree,

(c) the leading coefficient C_{20} has been allowed to be weighted separately,

and

(d) all the weights have been given a great amount of flexibility, in that magnitudes of the standard errors have been allowed to vary.

Without such changes, adjustments of real data could sometimes produce highly unrealistic corrections to the coefficients, especially if the data are insufficient.

A feature newly introduced in SAGG allows for predictions, after the adjustment, of geoid undulations and/or gravity anomalies at prescribed locations. Usually, these predictions are carried out in a geographic grid (e.g. a $5^0 \times 5^0$ grid), which can serve in drawing contour maps of these quantities. The predictions can be computed after a separate adjustment of either kind, or after a combined adjustment. A part of this feature is the associated variance-covariance propagation. Accordingly, contour maps can be drawn also for a-posteriori standard errors in geoid undulations and/or gravity anomalies.

The new version of SAGG has been used in a combined adjustment of real satellite altimetry data and real gravity anomaly data, in conjunction with a (14, 14) geopotential model. A comparison of internal precision has

demonstrated the beneficial effect of adding altimetry data to the existing body of gravity anomaly data. However, adding this relatively new and completely independent source to a more traditional type of data in a combined adjustment is beneficial also for other reasons, such as increasing the external reliability of the results. An important fact that has been confirmed during the real data reductions -- and not only in some computer simulations -- is that highly accurate reference orbits are not a stringent requirement in the short arc mode of satellite altimetry.

The recovered geoid over most of the globe shows good agreement with the gravimetric geoids obtained from independent sources. This is especially true of the areas covered by the GEOS-3 satellite when compared with the earlier reported results of the AFGL computer program SARRA (Short Arc Reduction of Radar Altimetry). It is noteworthy that the mathematical model used in SARRA is completely different from the one in SAGG.

In the present combined adjustment, only a part of the existing satellite altimetry and gravity anomaly data has been utilized. This indicates that the global geoid recovery with a standard error better than one meter could be achieved without undue difficulty. Such a standard error is interpreted as a precision indicator of the long-wavelength geoidal features, described by the truncated spherical harmonic model.

In a separate part of the current effort, a theoretical study has been performed in comparing the spherical harmonic representation of the potential with the spheroidal harmonic representation. The goal of this analysis has been to determine whether the latter representation is superior to the former.

However, such a superiority could not be proven. Furthermore, the practical difficulties connected with an eventual adjustment process in terms of spheroidal harmonics could often reach prohibitive levels. The outcome of this study has indicated that truncated expressions in any coordinate system other than the spherical (e.g., in a spheroidal coordinate system) would be essentially equivalent to using spherical harmonics, and they would be far less convenient.

The last, independent part of the present effort has been devoted to an analysis of the point mass technique. The purpose of introducing the point mass parameters into the adjustment is to add fine structure to a geopotential model that is based on spherical harmonic coefficients. The main attractive attributes of such an approach are its economy and its flexibility. In particular, in areas where an accurate geoid determination is not required or cannot be accomplished due to the lack of data, no point masses would be introduced; the geoid in such areas would accordingly be described by potential coefficients alone. This means that on the whole, the number of parameters could be kept relatively small. The point masses would be employed in areas where, for various reasons, a more detailed knowledge of geoid features may be required.

The study related to the point masses has been based to a large extent on the development in [Needham, 1970]. However, several deviations from this approach have been suggested. One such suggestion implies that the spherical harmonic coefficients could be adjusted separately, followed by a new adjustment exclusively in terms of the point mass paramaters. The new

"observations" would be the residuals of the previous adjustment taken with the opposite sign. Ideally, the final results would be equivalent to those that would be obtained if a simultaneous adjustment of the potential coefficients and point mass parameters were actually carried out; in practice, such a simultaneous adjustment would be economically prohibitive in most cases.

Another important recommendation concerns the number of constraints. In theory, the constraints have been separated into a local type and a global type. The local constraints are designed to keep the mass of the geoid and its volume unchanged in certain volume regions taken one at a time, in spite of the introduction of the point masses. The suggested number of local constraints in this analysis is about one half of the number used in [Needham, 1970]. A quite interesting conclusion has been reached regarding the global constraints. It appears that if the local constraints are properly applied, no global constraints of any kind are required in the point mass adjustment model. In conclusion, the point mass technique has an immediate appeal not only for its relative simplicity and its power as a practical tool in representing detailed geoidal features, but also because it is founded on a plausible physical concept.

APPENDIX 1

GEOMETRY OF SATELLITE ALTIMETRY IN PRACTICE

The geometry of the satellite altimetry adjustment model has been described in [Brown, 1973], pages 8 and 9. Although in reality the measurements are made along the direction perpendicular to the geoid (or more precisely, to the ocean surface), the direction perpendicular to the reference ellipsoid is considered to be a good approximation. This does not mean that the distance to the geoid is replaced by a distance to the ellipsoid, but that the direction associated with the distance to the geoid is rendered computationally accessible; the resulting errors have been deemed completely negligible in the above reference.

In what follows we shall rely heavily on the geometry of an ellipsoid. The final formula representing the mathematical model will be presented in two forms, the first suitable for numerical computations and the second suitable especially for theoretical analyses. The topic will be presented in several sections and subsections containing mostly straightforward derivations with a minimum amount of text.

A1.1 Preliminary Relations for an Ellipsoid

An ellipse (or a meridian section of an ellipsoid of revolution) with a few relevant quantities is depicted in Figure A1.1, where the introduced notations are self-explanatory. It is worth mentioning that the quantities associated with points on the ellipsoid will be primed. From

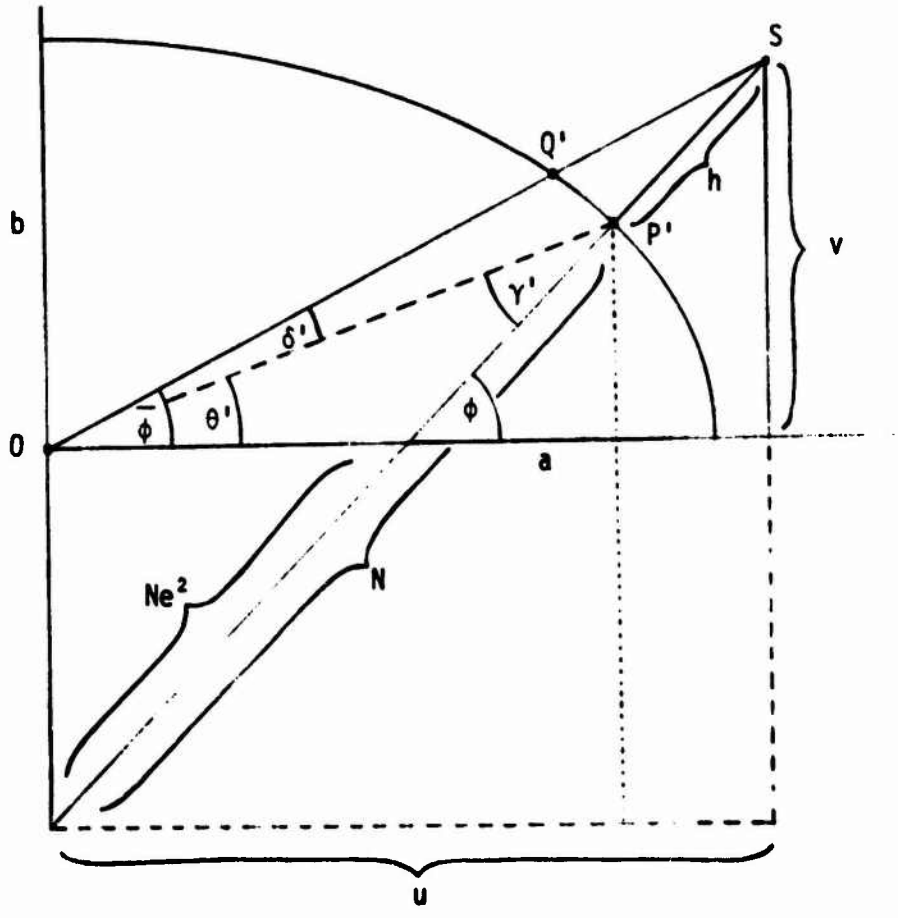


FIGURE A1.1

Meridian section of an ellipsoid of revolution

the figure we deduce the following relations:

$$u = (N+h)\cos\phi, \qquad \cos\phi = u/(N+h), \tag{A1.1}$$

$$v = \left[N(1-e^2)+h\right]\sin\phi, \qquad \sin\phi = v/\left[N(1-e^2)+h\right], \tag{A1.2}$$

$$tg\phi = (v/u)(N+h)/\left[N(1-e^2)+h\right]; \tag{A1.3}$$

$$tg\bar{\phi} = v/u; \tag{A1.4}$$

hence

$$tg\phi = tg\bar{\phi}\ (N+h)/\left[N(1-e^2)+h\right], \tag{A1.5}$$

$$tg\bar{\phi} = tg\phi\left[N(1-e^2)+h\right]/(N+h); \tag{A1.5'}$$

$$tg\theta' = \left[N(1-e^2)\sin\phi\right]/(N\cos\phi) \equiv (1-e^2)tg\phi. \tag{A1.6}$$

The last equation of course corresponds to (A1.5') with h=0 and $\bar{\phi}=\theta'$. From (A1.5') and (A1.6), we have

$$tg\bar{\phi} = \left[tg\theta'/(1-e^2)\right]\left[N(1-e^2)+h\right]/(N+h). \tag{A1.7}$$

Consider an ellipsoid of revolution centered at the origin of Cartesian coordinates and let its minor axis coincide with the z-axis. With the longitude (λ) measured in the positive sense (counterclockwise, the coordinate system being right-handed), the point S has the coordinates:

$$\left.\begin{aligned} X &= u\cos\lambda, \\ Y &= u\sin\lambda, \end{aligned}\right\} \quad u = (X^2+Y^2)^{\frac{1}{2}}$$

$$Z = v,$$

or

$$\left.\begin{aligned} X &= (N+h)\cos\phi\,\cos\lambda, \\ Y &= (N+h)\cos\phi\,\sin\lambda, \\ Z &= [N(1-e^2)+h]\,\sin\phi. \end{aligned}\right\} \qquad \text{(A1.8)}$$

From (A1.8) with the aid of (A1.3) and (A1.1) or (A1.2), the geodetic coordinates (ϕ,λ,h) are deduced from the Cartesian coordinates (X,Y,Z) as follows:

$$\tan\lambda = Y/X, \qquad \text{(A1.9)}$$

$$\tan\phi = [Z/(X^2+Y^2)^{\frac{1}{2}}]\,(N+h)/[N(1-e^2)+h], \qquad \text{(A1.10)}$$

$$h = (X^2+Y^2)^{\frac{1}{2}}/\cos\phi - N, \qquad \text{(A1.11)}$$

or

$$h = Z/\sin\phi - N(1-e^2), \qquad \text{(A1.11')}$$

where

$$N = a/(1-e^2\,\sin^2\phi)^{\frac{1}{2}}. \qquad \text{(A1.12)}$$

The equations (A1.10) through (A1.12) may serve in an iterative solution for ϕ and h; as a starting value, $Z/(X^2+Y^2)^{\frac{1}{2}}$ could be used for $\tan\phi$.

A1.2 Analytic Expressions for γ' and δ'

From Figure A1.1, the basic relations for these two quantities are

$$\gamma' = \phi - \theta' \quad ,$$

$$\delta' = \bar{\phi} - \theta' \quad .$$

With the aid of (A1.6), we have

$$\begin{aligned} e^2 \, tg\phi &= tg\phi - tg\theta' \\ &= tg(\phi-\theta')(1+tg\phi tg\theta') \\ &\simeq tg\gamma'(1+tg^2\phi), \end{aligned}$$

and, neglecting the terms with e^4, etc.,

$$\begin{aligned} tg\gamma' &\simeq e^2 \sin\phi \cos\phi, \\ \gamma' &\simeq e^2 \sin\phi \cos\phi = \tfrac{1}{2} e^2 \sin 2\phi. \end{aligned} \qquad \text{(A1.13)}$$

If $e^2 \simeq 0.006695$ is accepted (a rounded value corresponding to the mean earth ellipsoid in the GRS 1967 system), we can write

$$\gamma' \leq \tfrac{1}{2}e^2 \simeq 0.003347 \simeq 11'30''. \qquad \text{(A1.13')}$$

Similarly, from (A1.7) we have

$$tg\bar{\phi} \simeq tg\theta' \, (N+h+he^2)/(N+h),$$

$$\begin{aligned} tg\theta' \, he^2/(N+h) &\simeq tg\bar{\phi} - tg\theta' \\ &\simeq tg(\bar{\phi}-\theta')(1+tg\bar{\phi}tg\theta') \\ &\simeq tg\delta'(1+tg^2\theta'); \end{aligned}$$

$$tg\delta' \simeq [he^2/(N+h)]\sin\theta' \cos\theta' = \tfrac{1}{2}e^2 \sin 2\theta' \, h/(N+h),$$

$$\delta' \simeq \tfrac{1}{2}e^2 \sin 2\theta' \, h/(N+h) \qquad \text{(A1.14)}$$

or, if we replace N by r_0 (r_0 in this analysis may safely be considered as the mean earth's radius), $r_0 \simeq 6371$ km,

$$\delta' \simeq \tfrac{1}{2}e^2 \sin 2\theta' \, h/(r_0+h). \qquad \text{(A1.14')}$$

Upon using the above values for e^2 and r_0, and upon further assuming the realistic value $h \simeq 1000$ km associated with satellite altimetry, one obtains

$$\delta' \leq 0.0004541 \simeq 1'34". \qquad \text{(A1.14")}$$

Finally, the distance P'Q' is determined as

$$P'Q' \simeq r_0 \, \delta' \quad , \qquad \text{(A1.15)}$$

$$P'Q' \leq 2.89 \text{ km} . \qquad \text{(A1.15')}$$

A1.3 Geometry of Satellite Altimetry

A1.3.1 Basic Relations

A meridian section of the reference ellipsoid containing the altimetry measurement (H), assumed to be made perpendicular to the ellipsoid, is depicted in Figure A1.2. The notations are again self-explanatory.

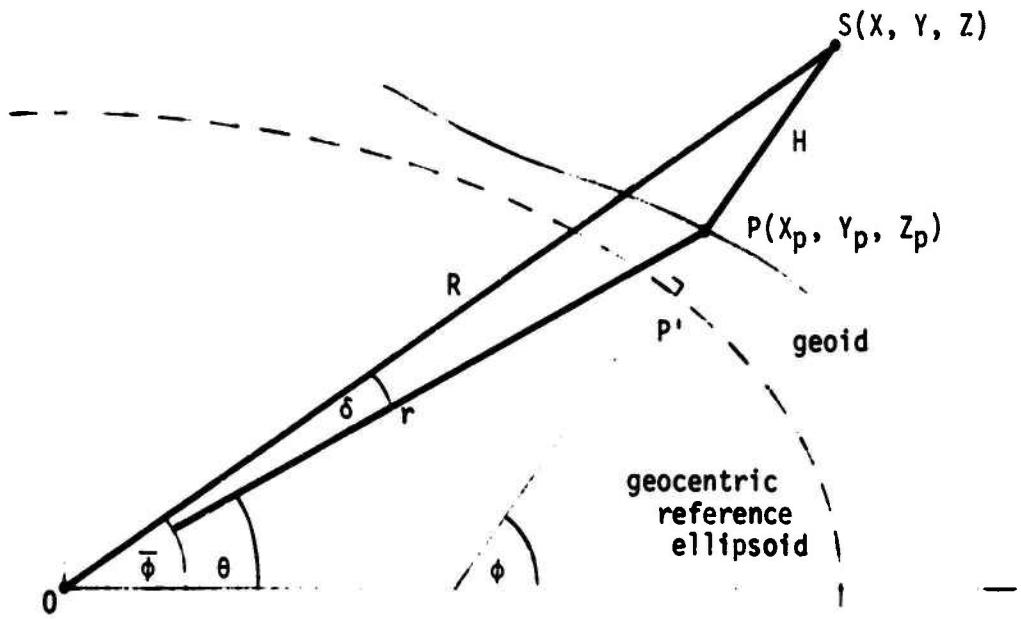


FIGURE A1.2

Triangle OPS of satellite altimetry and its relation to the reference ellipsoid

The model equation is seen to be

$$H = R - r + d \quad , \tag{A1.16}$$

where d depends on δ. Near the poles or the equator, $\delta \simeq 0$ and $H \simeq R - r$; the omission of the correction term d is the most dangerous in mid-latitudes since δ is maximum at $\theta \simeq 45^0$.

The distance R is computed from the satellite coordinates (X, Y, Z), while for a given θ, r may be computed from the earth's potential coefficients. To find θ, the coordinates of P are computed as

$$X_p = X - H \cos \phi \cos \lambda \quad ,$$
$$Y_p = Y - H \cos \phi \sin \lambda \quad ,$$
$$Z_p = Z - H \sin \phi \quad ,$$

where ϕ is obtained through the iterative process depicted in (A1.10) to (A1.12), and λ is given by (A1.9). As a by-product, the geoid undulation at P could be computed from

$$P'P = h - H \quad .$$

We thus have

$$\text{tg } \theta = Z_p / (X_p^2 + Y_p^2)^{\frac{1}{2}} \quad . \tag{A1.17}$$

Knowing θ, we can compute

$$\delta = \overline{\phi} - \theta \quad , \tag{A1.17'}$$

where

$$\text{tg } \overline{\phi} = Z / (X^2 + Y^2)^{\frac{1}{2}} \quad . \tag{A1.17"}$$

If we can show that $\delta = \delta'$ and thus $\theta = \theta'$ are valid with sufficient accuracy, our task of expressing d analytically will be greatly facilitated since δ' has been already given, in a suitable form, by (A1.14'). If the point P' were moved along P'S by the amount of P'P in Figure A1.1, the change in θ' would be

$$d\theta' \simeq (P'P)\,\gamma'/r_0 \quad .$$

Upon using (A1.13') and upon considering P'P as unrealistically high as 200m, we conclude that

$$d\theta' \leq 0.02" \quad .$$

Thus we can indeed write

$$\theta' = \theta \quad , \qquad \delta = \delta' \quad , \tag{A1.18}$$

and, from (A1.14'),

$$\delta \simeq \tfrac{1}{2}\, e^2 \left[H/(r_0 + H)\right] \sin 2\theta \quad , \tag{A1.19}$$

where use was made of

$$h / (r_0 + h) \simeq H / (r_0 + H) \quad ;$$

this last approximation would affect the fifth significant digit at the most. Similar to (A1.14"), we also have

$$\delta \leq 0.0004541 \simeq 1'34" \quad . \tag{A1.19'}$$

A1.3.2 Analytic Expression for d

The pertinent part of Figure A1.2 is reproduced and specialized for the present task as Figure A1.3. The notations are again self-explanatory and the straightforward derivations are presented below:

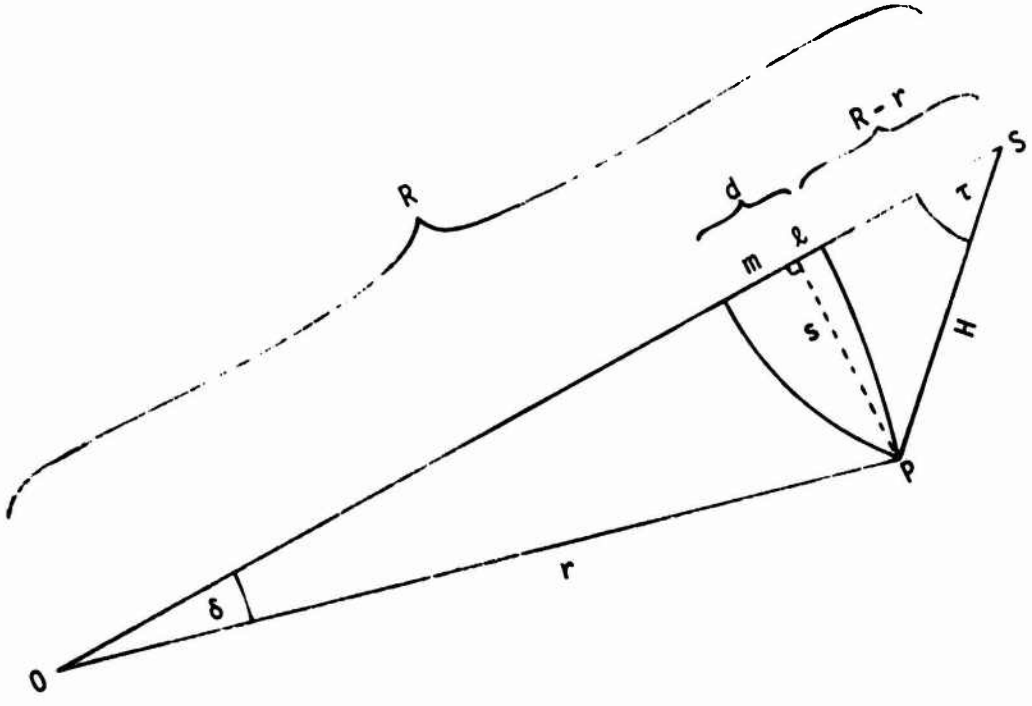


FIGURE A1.3

Traingle OPS and the quantities needed in expressing d

$$\tau \simeq s/H , \qquad \delta \simeq s/r ,$$

$$\tau \simeq \delta r/H \quad ;$$

$$m = H (1 - \cos \tau)$$

$$\simeq \tfrac{1}{2} H \tau^2$$

$$\simeq \tfrac{1}{2} \delta^2 r^2/H \quad ,$$

$$\ell = r (1 - \cos \delta) \simeq \tfrac{1}{2} r \delta^2 \quad ,$$

$$d = m + \ell \simeq \tfrac{1}{2} r \delta^2 (1 + r/H) ,$$

$$d \simeq \tfrac{1}{2} r \delta^2 (r + H)/H \quad . \qquad \text{(A1.20)}$$

Upon replacing r by r_0 (which could hardly affect even the third significant digit), one can write

$$d \simeq \tfrac{1}{2} r_0 \delta^2 (r_0 + H)/H \quad . \qquad \text{(A1.20')}$$

To express d analytically, we make use of (A1.19) in (A1.20'), which yields

$$d \simeq (1/8) e^4 r_0 [H/(r_0 + H)] \sin^2 2\theta \quad . \qquad \text{(A1.21)}$$

When the usual values of e^2 and r_0 are substituted for in (A1.21), one obtains

$$d \simeq 35.9 \text{ m } [H/(r_0 + H)] \sin^2 2\theta \quad ; \qquad \text{(A1.21')}$$

considering again H ≃ 1000 km, this is

$$d \simeq 4.9\ m \sin^2 2\theta\ , \qquad \text{(A1.21")}$$

so that

$$d \leq 4.9m\ .$$

A1.3.3 Summary of Formulas

The given constants needed in these computations comprise a and e^2 of the geocentric reference ellipsoid, r_o (here the value of the earth's mean radius), satellite coordinates X, Y, Z, and the value of satellite altimetry, H. The computations proceed as follows:

1) Compute ϕ, λ:

$$tg\lambda = Y/X\ ;$$

$$\left.\begin{aligned} tg\phi &= [Z/(X^2+Y^2)^{\frac{1}{2}}](N+h)/[N(1-e^2)+h], \\ N &= a/(1-e^2\sin^2\phi)^{\frac{1}{2}}\ , \\ h &= (X^2+Y^2)^{\frac{1}{2}}/\cos\phi - N\ ,\quad \phi \neq \tfrac{1}{2}\pi\ . \end{aligned}\right\} \text{iterations}$$

2) Compute θ and r:

$$X_p = X - H\cos\phi\cos\lambda\ ,$$

$$Y_p = Y - H\cos\phi\sin\lambda\ ,$$

$$Z_p = Z - H\sin\phi\ ;$$

$$tg\theta = Z_p/(X_p^2+Y_p^2)^{\frac{1}{2}}\ .$$

For the given θ, λ,

$$r = r \quad \text{(potential coefficients; } \theta, \lambda).$$

3) Compute δ:

$$\delta = \overline{\phi} - \theta \ ,$$

where $\overline{\phi}$ is determined from

$$tg\overline{\phi} = Z / (X^2 + Y^2)^{\frac{1}{2}} \ .$$

The analytical expression for δ is not used in actual computations, but if needed, it would have the form

$$\delta \simeq \tfrac{1}{2} e^2 [H / (r_o + H)] \sin 2\theta \ .$$

4) Compute d:

$$d = \tfrac{1}{2} r \delta^2 (r + H) / H \ .$$

The analytical expression for d could also be used; it reads

$$d \simeq (1/8) e^4 r_o \left[H / (r_o + H) \right] \sin^2 2\theta \ ;$$

to some approximation,

$$d \simeq 4.9 \text{ m } \sin^2 2\theta \text{ ... per 1000 km in H.}$$

5) The "misclosure" of satellite altimetry follows from (A1.16) as

$$\text{"misclosure"} = R - r + d - H \ ,$$

where

R = R (initial X, Y, Z as computed from intial orbital parameters),

r = r (initial potential coefficients; θ, λ),

H = the observed value of altimetry,

d = the above computed correction term.

APPENDIX 2

DERIVATION OF AN ACCURATE FORMULA EXPRESSING THE DIFFERENCE BETWEEN THE NORMAL GRAVITY AND ITS RADIAL COMPONENT

A2.1 Introduction

We propose to derive a formula giving the value of $\gamma-\gamma_r$ reliable to 0.001 mgal, where γ is the normal gravity and γ_r is the radial component of normal gravity, both associated with a given level ellipsoid. The value $\gamma-\gamma_r$ will be a function of the ellipsoidal parameters and of the geocentric latitude ($\bar{\phi}$). The formula will be derived in a straightforward manner, built on the notion of the potential and its first derivatives in the normal gravity field (i.e., in the gravity field of the level ellipsoid, see [Heiskanen and Moritz, 1967], Section 2-7). Attention will be paid to preserving the necessary number of significant digits while simplifying the functional relationships wherever possible. Since the largest possible value of $\gamma-\gamma_r$ is slightly over 5.5 mgal (see e.g. [Blaha, 1975], equation 38), the accuracy to be maintained in the derivations is characterized by at least four reliable significant digits. The procedure followed will be a significantly refined version of Section 3.1 of [Blaha, 1975]; in particular, the validity of the formula contained therein will be extended by approximately two significant digits.

A2.2 Basic Relations

In addition to γ_r, we shall be dealing with $\gamma_{\bar{\phi}}$ which is the other nonzero component of normal gravity in spherical coordinates; we have

$$\gamma = (\gamma_r^2 + \gamma_{\bar{\phi}}^2)^{\frac{1}{2}}, \tag{A2.1a}$$

where

$$\gamma_r = -\partial U/\partial r, \tag{A2.1b}$$

$$\gamma_{\bar{\phi}} = (1/r)\, \partial U/\partial \bar{\phi}, \tag{A2.1c}$$

while $\gamma_\lambda = 0$; U represents the normal potential and r is the radial distance from the coordinate origin to the point where γ is being computed. From (A2.1a) it is deduced that

$$\gamma = \gamma_r [1 + (\gamma_{\bar{\phi}}/\gamma)^2]^{\frac{1}{2}},$$

$$\gamma_r = \gamma\, [1 + (\gamma_{\bar{\phi}}/\gamma_r)^2]^{-\frac{1}{2}},$$

$$\gamma - \gamma_r = \tfrac{1}{2}(\gamma_{\bar{\phi}}/\gamma_r)^2 [1-(3/4)(\gamma_{\bar{\phi}}/\gamma_r)^2 + \ldots]\gamma.$$

Since the second term inside the brackets reaches at most 0.0000086 (see equation 37 of [Blaha, 1975]), it can be safely neglected and the expression to be further developed is

$$\gamma - \gamma_r = \tfrac{1}{2}\, (\gamma_{\bar{\phi}}/\gamma_r)^2 \gamma. \tag{A2.2}$$

The standard formula for the normal potential at a point in space (the point rotates with the reference ellipsoid) follows from [Heiskanen and Moritz, 1967], page 73, as

$$U = (kM/r)[1-J_2(a/r)^2\, P_2(\sin\bar{\phi}) - J_4(a/r)^4\, P_4(\sin\bar{\phi}) - \ldots] + \tfrac{1}{2}\, \omega^2 r^2 \cos^2\bar{\phi}, \tag{A2.3}$$

where k is the gravitational constant, M and ω are the mass and the rotational velocity of the reference level ellipsoid, J_2, J_4, etc., are the coefficients related to the flattening of this ellipsoid, a is its semi-major axis, $P_{2n}(\sin\bar{\phi})$ are the Legendre functions, and $r, \bar{\phi}$ are two spherical coordinates -- the third being λ -- of the point in question (here considered outside the ellipsoid's mass). The potential on the ellipsoid is $U=U_0$. The ellipsoid is centered at the coordinate origin and its axis of rotation coincides with the Cartesian z- axis. The coefficients J are also identified as

$$J_{2n} \equiv -C^*_{2n,0}.$$

We mention that if the four parameters M, ω, U_0, and J_2 are shared between the level ellipsoid and the actual earth, the uniquely determined ellipsoid is then the mean earth ellipsoid.

In addition to the four independent parameters just mentioned from which any other ellipsoidal parameter may be derived, for convenience we define

$$r_0^* = kM/U_0, \tag{A2.4}$$

where the parameter r_0^* is the radius of a homogeneous nonrotating sphere whose potential is U_0. Upon specializing (A2.3) for $U=U_0$ and multiplying its both sides by $r\, r_0^*/(kM)$, one obtains

$$r = r_0^*\left[1-J_2(a/r)^2\, P_2(\sin\bar{\phi}) - J_4(a/r)^4\, P_4(\sin\bar{\phi}) - \ldots + \tfrac{1}{2}\,\omega^2 r^3 \cos^2\bar{\phi}/(kM)\right], \tag{A2.5}$$

where r refers to points on the level ellipsoid sharing the given $\bar{\phi}$. After simplifications, this equation will serve in expressing $\gamma_{\bar{\phi}}/\gamma_r$ in terms of the ellipsoidal parameters (and $\bar{\phi}$).

A2.3 Simplifications in the Rigorous Mathematical Model

In agreement with (A2.1b) and (A2.1c), from (A2.3) we express, reliable to four significant digits:

$$\gamma_r = (kM/r^2)\left[1-3\, J_2(a/r)^2\, P_2(\sin\bar{\phi}) - \omega^2 r^3 \cos^2\bar{\phi}/(kM)\right], \qquad \text{(A2.6a)}$$

$$\gamma_{\bar{\phi}} = (kM/r^2)\left[-J_2(a/r)^2\, dP_2(\sin\bar{\phi})/d\bar{\phi} - J_4(a/r)^4\, dP_4(\sin\bar{\phi})/d\bar{\phi} - \omega^2 r^3 \sin\bar{\phi}\cos\bar{\phi}/(kM)\right]. \qquad \text{(A2.6b)}$$

The magnitude of the term containing J_4 in (A2.6a) would not exceed 0.000012 if it were indeed present; this, compared with unity, is negligible for our purpose. Similarly, the magnitude of the term containing J_6 in (A2.6b) would not exceed $0.000000061\ |\sin\bar{\phi}\cos\bar{\phi}|$, which is negligible when compared with the term containing J_2 (its value is approximately $-.00324\ \sin\bar{\phi}\cos\bar{\phi}$) considered together with the term containing ω^2 (this term amounts to approximately $-.00346\ \sin\bar{\phi}\cos\bar{\phi}$). The numerical values that enable us to make these and similar simplifications are the following:

$$J_2 \approx 1.08 \times 10^{-3}, \quad J_4 \approx -2.37 \times 10^{-6}, \quad J_6 \approx 6.1 \times 10^{-9};$$
$$kM/r^2 \approx 980 \text{ gal}, \quad \omega^2 r \approx 3.39 \text{ gal}, \quad \omega^2 r^3/(kM) \approx 0.00346;$$

$$P_2(\sin\bar{\phi}) = \tfrac{1}{2}(3\sin^2\bar{\phi} - 1) \le 1,$$
$$P_4(\sin\bar{\phi}) = (1/8)(35\sin^4\bar{\phi} - 30\sin^2\bar{\phi} + 3) \le 1,$$
$$P_6(\sin\bar{\phi}) = (1/16)(231\sin^6\bar{\phi} - 315\sin^4\bar{\phi} + 105\sin^2\bar{\phi} - 5) \le 1;$$

$$dP_2(\sin\bar{\phi})/d\bar{\phi} = 3\sin\bar{\phi}\cos\bar{\phi},$$
$$dP_4(\sin\bar{\phi})/d\bar{\phi} = (5/2)(7\sin^2\bar{\phi} - 3)\sin\bar{\phi}\cos\bar{\phi} \le 10\,|\sin\bar{\phi}\cos\bar{\phi}|,$$
$$dP_6(\sin\bar{\phi})/d\bar{\phi} = (3/8)(231\sin^4\bar{\phi} - 210\sin^2\bar{\phi} + 35)\sin\bar{\phi}\cos\bar{\phi} \leqslant 56\,|\sin\bar{\phi}\cos\bar{\phi}|.$$

Upon dividing (A2.6b) by (A2.6a) and expressing the Legendre functions and their first derivative with respect to $\bar{\phi}$ explicitly, we obtain

$$\gamma_{\bar{\phi}}/\gamma_r = -[3J_2(a/r)^2 + (5/2)\, J_4(a/r)^4(7\sin^2\bar{\phi} - 3) + \omega^2 r^3/(kM)]$$
$$\times [1-(3/2)\, J_2(a/r)^2\, (3\sin^2\bar{\phi} - 1) - \omega^2 r^3 \cos^2\bar{\phi}/(kM)]^{-1} \sin\bar{\phi}\cos\bar{\phi}.$$

In considering the required precision, the inversion of the quantity inside the second pair of brackets may be accommodated by changing the signs of the two terms following unity. When the subsequent multiplication of this result by the quantity inside the first pair of brackets is carried out, the two largest terms are $3J_2(a/r)^2 \simeq 0.00324$ and $\omega^2 r^3/(kM) \simeq 0.00346$. It follows that the number of decimals to be retained is six and, accordingly, that two terms resulting from this multiplication may be neglected (they involve the second term in the first pair of brackets). We further replace $3\sin^2\bar{\phi} - 1 + 2\cos^2\bar{\phi}$ by $1 + \sin^2\bar{\phi}$ and $\cos^2\bar{\phi}$ by $1 - \sin^2\bar{\phi}$; satisfying the required precision, we may finally write

(A2.7)

$$\gamma_{\bar{\phi}}/\gamma_r = -\{3J_2(a/r)^2 + \omega^2 r^3/(kM) + (5/2)\, J_4(7\sin^2\bar{\phi} - 3) + (9/2)\, J_2^2(3\sin^2\bar{\phi} - 1)$$
$$+ (3/2)\, J_2[\omega^2 a^3/(kM)](1+\sin^2\bar{\phi}) + [\omega^2 a^3/(kM)]^2(1-\sin^2\bar{\phi})\}\sin\bar{\phi}\cos\bar{\phi},$$

where, in all terms except the first two, r was replaced by a. Since these modified terms are two orders of magnitude smaller than the first two terms, they should be reliable to only two significant digits and the simplification is amply justified.

The next task consists in expressing the first two terms on the right-hand side of (A2.7) in terms of "a" in such a way that the number of four reliable significant digits is preserved. This will be done mathematically by stipulating that r refers to a point on the reference ellipsoid. However, it is clear that the final formula will not be overly sensitive to r since it will hold, to all the significant digits required, even if r is increased by a few hundred meters, i.e., even if the point for which $\gamma-\gamma_r$ is being determined is located a few hundred meters above the reference ellipsoid. With sufficient accuracy, from (A2.5) we have

$$r = r_0^* \left[1 - J_2 P_2(\sin\bar{\phi}) + \tfrac{1}{2}\omega^2 a^3 \cos^2\bar{\phi}/(kM)\right].$$

If we write

$$r = a + da,$$

it then follows that

$$da = r_0^* - a - r_0^* J_2 P_2(\sin\bar{\phi}) + \tfrac{1}{2} r_0^* \omega^2 a^3 \cos^2\bar{\phi}/(kM).$$

Since

$$r^{-2} = a^{-2} - 2a^{-3}\, da,$$

we have

$$(a/r)^2 = a - 2da/a$$

and

$$(a/r)^2 = 1 - 2\left[(r_0^*-a)/a - \tfrac{1}{2} J_2(3\sin^2\bar{\phi}-1) + \tfrac{1}{2}t\cos^2\bar{\phi}\right], \qquad \text{(A2.8a)}$$

where

$$t \equiv \omega^2 a^3/(kM) \qquad \text{(A2.8b)}$$

and where, inside the brackets, $P_2(\sin\bar{\phi})$ was expressed explicitly and r_0^*/a was set to unity; this simplification is justified upon realizing that only two reliable significant digits are needed in each of the three terms, the leading (first) term being approximately -0.00227. In much the same way, from

$$r^3 = a^3(1 + 3da/a)$$

one obtains (A2.8c)

$$\omega^2 r^3/(kM) = t + 3t\left[(r_0^*-a)/a - \tfrac{1}{2} J_2(3\sin^2\bar{\phi}-1) + \tfrac{1}{2}t\cos^2\bar{\phi}\right].$$

The substitution of equations (A2.8) into (A2.7) yields

$$\gamma_{\bar{\phi}}/\gamma_r = -\{3J_2+t-3(2J_2-t)\left[(r_0^*-a)/a - \tfrac{1}{2}J_2(3\sin^2\bar{\psi}-1) + \tfrac{1}{2}t(1-\sin^2\bar{\phi})\right]$$
$$+(5/2)J_4(7\sin^2\bar{\phi}-3) + (9/2)J_2^2(3\sin^2\bar{\phi}-1) + (3/2)J_2 t(1+\sin^2\bar{\phi}) + t^2(1-\sin^2\bar{\phi})\}$$
$$\times \sin\bar{\phi}\cos\bar{\phi},$$

which after some cancellation becomes (A2.9)

$$\gamma_{\bar{\phi}}/\gamma_r = -\{(3J_2+t) - 3(2J_2-t)(r_0^*-a)/a - (15/2)J_2^2 - (15/2)J_4 + (5/2)t^2$$
$$+\sin^2\bar{\phi}\left[(45/2)J_2^2 + (35/2)J_4 - (5/2)t^2\right]\}\sin\bar{\phi}\cos\bar{\phi}.$$

The leading term $(3J_2+t)$ in (A2.9) is approximately 0.00670, while each of the following four terms is at least two orders of magnitude smaller (the first three of them nearly cancel out); the sum of these four terms is over 200 times smaller than the leading term. A similar conclusion applies for all the three terms multiplied by $\sin^2\bar{\phi}$, whose sum in absolute value is about 150 times smaller than the leading term. It thus follows that when squaring (A2.9), only those terms which contain $(3J_2+t)$ as a factor will be kept; in

particular,

$$(\gamma_{\bar{\phi}}/\gamma_r)^2 = \Big\{\big[(3J_2+t) - 6(2J_2-t)(r_o^*-a)/a - 15J_2^2 - 15J_4 + 5t^2\big](3J_2+t) + \sin^2\bar{\phi}(45J_2^2 + 35J_4 - 5t^2)(3J_2+t)\Big\}(1/4)\sin^2 2\bar{\phi}.$$

If this last equation is now multiplied by $\frac{1}{2}\gamma$, we obtain the desired formula for $\gamma-\gamma_r$ (see equation A2.2).

A2.4 Final Formula and Its Simplified Versions

In view of the outcome of the last section, it is now possible to present the final formula for a quite arbitrary level ellipsoid as follows:

$$\gamma-\gamma_r \equiv f(a,J_2,\ldots,\bar{\phi}) = (A_1-A_2 \sin^2\bar{\phi})\,\gamma \sin^2 2\bar{\phi}, \qquad \text{(A2.10)}$$

where

$$A_1 = A_3[A_4 + 5t^2 - 15J_4 - 15J_2^2 - 6(2J_2-t)\,A_5], \qquad \text{(A2.11a)}$$

$$A_2 = A_3\,(5t^2 - 35J_4 - 45J_2^2); \qquad \text{(A2.11b)}$$

$$A_3 = A_4/8, \qquad \text{(A2.11c)}$$

$$A_4 = 3J_2+t, \qquad \text{(A2.11d)}$$

$$A_5 = (r_0^*-a)/a, \qquad \text{(A2.11e)}$$

$$t = \omega^2 a^3/(kM). \qquad \text{(A2.11f)}$$

For a given ellipsoid, the values of A_1 and A_2 are computed once and for all. The normal gravity (γ) may be computed by the rigorous formula of Somigliana,

$$\gamma = (a\gamma_e \cos^2\phi + b\gamma_p \sin^2\phi)/(a^2 \cos^2\phi + b^2 \sin^2\phi)^{\frac{1}{2}},$$

where γ_e and γ_p are the normal gravity at the equator and at the poles, respectively, b is the semi-minor axis of the reference ellipsoid, and ϕ is the geodetic latitude. The quantities a, b, and $\phi,\bar{\phi}$ are related through the eccentricity (e); in particular,

$$b = a(1-e^2)^{\frac{1}{2}},$$

$$\text{tg}\,\bar{\phi} = (1-e^2)\text{tg}\phi.$$

We shall next express the accurate formula (A2.10) numerically for a given ellipsoid and shall present its two simplified versions.

If the reference level ellipsoid is the mean earth ellipsoid in the Geodetic Reference System 1967 (GRS 1967), the following rounded values may be used for computing the parameters of equations (A2.11) with sufficient accuracy:

$$r_0^* = 6363696 \text{ m},$$

$$a = 6378160 \text{ m},$$

$$\omega = .72921\times10^{-4} \text{ rad/sec},$$

$$kM = 3.9860\times10^{14} \text{ m}^3/\text{sec}^2,$$

$$J_2 = 1.0827\times10^{-3},$$

$$J_4 = -2.3713\times10^{-6};$$

this yields

$$A_4 = 0.0067095,$$

$$A_3 = 0.00083868,$$

$$A_5 = -0.0022677,$$

$$t = 0.0034614,$$

and thus

$$A_1 = 5.6777\times10^{-6},$$

$$A_2 = 0.0756\times10^{-6}.$$

Accordingly, the accurate formula (A2.10) in the GRS 1967 is (to four significant digits)

(A2.12)

$$\gamma-\gamma_r \equiv f(a,J_2,\ldots,\bar{\phi}) = (5.678 - 0.076 \sin^2\bar{\phi})\ \gamma\times10^{-6} \sin^2 2\bar{\phi}.$$

Equation (A2.12) may be simplified if γ is replaced by 979.8 gal, the mean value of gravity. If we consider that $\gamma_e \approx 978.03$ gal and $\gamma_p \approx 983.22$ gal, the error introduced in the formula below reaches at most 0.009 mgal around

$\phi = \pm 60^0$. We have thus arrived at a single, yet quite accurate expression for practical computations, namely

$$\gamma - \gamma_r \simeq (5.563 - 0.074 \sin^2\bar{\phi}) \sin^2 2\bar{\phi} \text{ mgal.} \qquad \text{(A2.13)}$$

A further simplification is obtained if $\sin^2\bar{\phi}$ is replaced by its mean value of $\frac{1}{2}$. The added error reaches at most 0.014 mgal for $\bar{\phi} = \pm 62.6^0$ and $\bar{\phi} = \pm 27.4^0$ (its sign associated with one pair of values $\bar{\phi}$ is the opposite of the sign associated with the other pair). We thus have

$$\gamma - \gamma_r \simeq 5.526 \sin^2 2\bar{\phi} \text{ mgal,} \qquad \text{(A2.14)}$$

whose total error is certainly smaller than 0.023 mgal. In fact, around $\bar{\phi} = \pm 60^0$ the errors in (A2.14) from both simplifications have opposite signs; near the equator they have the same sign, but the error due to the mean value of γ has a very small magnitude. A maximum error of 0.017 mgal (around $\bar{\phi} = \pm 30^0$) is probably close to the reality. If the normal gravity has been computed beforehand for other reasons, the application of the formula (A2.12) is easy and efficient indeed and it has the advantage of high accuracy. The same holds of course true with regard to the general formula (A2.10).

To assess the accuracy of the formula (A2.10), a computer comparison was made between this formula and the errorless results $\gamma - \gamma_r$ for a number of ellipsoidal locations, with $\bar{\phi}$ ranging from pole to pole in 1^0 intervals. In computing either γ_r or $\gamma_{\bar{\phi}}$ during this comparison, the terms with J_4 and J_6 have been included not only in the explicit form of equations (A2.6), but also in the intermediate task of computing r according to equation (A2.5). The GRS 1967 has served to provide a consistent set of the input parameters.

For the sake of completeness, we present all the pertinent formulas used in computing the errorless results:

$$\gamma = (\gamma_r^2 + \gamma_{\bar{\phi}}^2)^{\frac{1}{2}},$$

$$\gamma_r = (kM/r^2)[1-3J_2(a/r)^2P_2(\sin\bar{\phi})-5J_4(a/r)^4P_4(\sin\bar{\phi}) - 7J_6(a/r)^6P_6(\sin\bar{\phi})] \\ -\omega^2 r \cos^2\bar{\phi},$$

$$\gamma_{\bar{\phi}} = (kM/r^2)[-J_2(a/r)^2dP_2(\sin\bar{\phi})/d\bar{\phi}-J_4(a/r)^4dP_4(\sin\bar{\phi})/d\bar{\phi}-J_6(a/r)^6dP_6(\sin\bar{\phi})/d\bar{\phi}] \\ -\omega^2 r \sin\bar{\phi}\cos\bar{\phi},$$

where r can be computed either directly as

$$r = a/[1+e^2 \sin^2\bar{\phi}/(1-e^2)]^{\frac{1}{2}},$$

or, equivalently, through an iterative process as

$$r = r_0^* [1-J_2(a/r)^2P_2(\sin\bar{\phi}) - J_4(a/r)^4P_4(\sin\bar{\phi}) - J_6(a/r)^6P_6(\sin\bar{\phi})] \\ +\tfrac{1}{2}\omega^2 r_0^* r^3 \cos^2\bar{\phi}/(kM).$$

The computer program used in these tests was originally designed to compute the quantities related to the geoid, for instance g and its three orthogonal components. All the quantities above, related now to the ellipsoid, can be obtained simply by setting $C_{20} = -J_2$, $C_{40} = -J_4$, $C_{60} = -J_6$, and by suppressing all the other spherical harmonic potential coefficients. The outcome of these tests indicated that the formula (A2.10) is completely satisfactory, accurate to better than 0.001 mgal. All the discrepancies between the values computed by (A2.10) and the errorless results have been listed as 0.000 mgal.

REFERENCES

Balmino, G., "Representation of the Earth Potential by Buried Masses", pp. 121-124 in The Use of Artificial Satellites for Geodesy, Geophysical Monograph 15, American Geophysical Union, Washington, D.C., 1972.

Blaha, G., The Combination of Gravity and Satellite Altimetry Data for Determining the Geoid Surface. Report of DBA Systems, Inc.; AFCRL Report No. 75-0347, Air Force Cambridge Research Laboratories, Hanscom AFB, Massachusetts, 1975.

Blaha, G., The Least Squares Collocation from the Adjustment Point of View and Related Topics. Report of DBA Systems, Inc.; AFGL Scientific Report No. 76-0073, Air Force Geophysics Laboratory, Hanscom AFB, Massachusetts, 1976.

Blaha, G., "Refinement of the Short Arc Satellite Altimetry Adjustment Model". Paper published in Bulletin Geodesique, Vol. 51, No. 1, Bureau Central de l'Association Internationale de Geodesie, Paris, France, 1977.

Brown, D.C., Review of Current Geodetic Satellite Programs and Recommendations for Future Programs. Report for NASA Headquarters, Contract No. NASW-1469, June 1967.

Brown, D.C., Investigation of the Feasibility of a Short Arc Reduction of Satellite Altimetry for Determination of the Oceanic Geoid. Report No. AFCRL-TR-73-0520, Air Force Cambridge Research Laboratories (LW), Bedford, Massachusetts, 1973.

Brown, D.C., "A Test of Short Arc Versus Independent Point Positioning Using the Broadcast Ephemeris of Navy Navigational Satellites". Paper presented at the 1975 Spring Annual Meeting of the American Geophysical Union, Washington, D.C., June 1975.

Dragg, J.L., "Applications of Spherical Harmonics to Geodetic Problems", ACIC Reference Publication No. 15, Aeronautical Chart and Information Center, Second and Arsenal Streets, St. Louis 18, Missouri, February 1963.

Gopalapillai, S., Non-Global Recovery of Gravity Anomalies from a Combination of Terrestrial and Satellite Altimetry Data. Department of Geodetic Science, Report No. 210, The Ohio State University, Columbus, 1974.

Hadgigeorge, G. and J.E. Trotter, "Short Arc Reductions of GEOS-3 Altimetry Data". Paper presented of the AGU Fall Meeting, December 1976.

Heiskanen, W.A. and H. Moritz, Physical Geodesy. W.H. Freeman and Co., San Francisco, 1967.

Hotine, M., Mathematical Geodesy. Monogr. Ser., Vol. 2, Environ. Sci. Serv. Admin., Washington, D.C., 1969.

Moritz, H., "Fundamental Geodetic Constants", pp. 129-144 in Contributions of the Graz Group to the XVI General Assembly of IUGG/IAG in Grenoble 1975, edited by P. Meissl, H. Moritz, K. Rinner; Technical University at Graz, Institute of Physical Geodesy, Graz, Austria, 1975.

Morrison, F., "Validity of the Expansion for the Potential Near the Surface of the Earth". Paper presented at the Proceedings of the IV Symposium on Mathematical Geodesy, Trieste, Italy, 1969.

Needham, P.E., The Formation and Evaluation of Detailed Geopotential Models Based on Point Masses. Department of Geodetic Science, Report No. 149, The Ohio State University, Columbus, 1970.

Rapp, R.H., The Role of Gravity Data in Determination of the Gravitational Potential of the Earth. Department of Geodetic Science, Report No. 134, The Ohio State University, Columbus, 1970.

Rapp, R.H., Accuracy of Potential Coefficients Determinations from Satellite Altimetry and Terrestrial Gravity. Department of Geodetic Science, Report No. 166, The Ohio State University, Columbus, 1971.

Rapp, R.H., Geopotential Coefficient Behavior to High Degree and Geoid Information by Wavelength. Department of Geodetic Science, Report No. 180, The Ohio State University, Columbus, 1972.

Rapp, R.H., Improved Models for Anomaly Computations from Potential Coefficients. Department of Geodetic Science, Report No. 181, The Ohio State University, Columbus, 1972'.

Rapp, R.H., Procedures and Results Related to the Direct Determination of Gravity Anomalies from Satellite and Terrestrial Gravity Data. Department of Geodetic Science, Report No. 211, The Ohio State University, Columbus, 1974.